AD-A096 512 PENNSYLVANIA UNIV PHILADELPHIA F/G 5/8
WORKSHOP ON INTERACTIVE MAN-MACHINE DISCOURSE HELD AT THE UNIVE--ETC(U)
1980 A K JOSHI N00014-80-G-0054

UNCLASSIFIED NL

1 OF 1
AD A096512

END
DATE
FILMED
4-81
DTIC

AD A096512

LEVEL II

(12)

Report on the

Workshop on Interactive Man-Manchine Discourse

held at the University of Pennsylvania 17-20 Jun 1980

June 17, 18, 19, 20, 1980

Sponsored by ONR

under grant N00014-80-G-0054

(Identifying number: NR 049-458)

Principal Investigator: Aravind K. Joshi

81 1 12 138

This workshop was organized to discuss some critical issues in the design of interactive natural language system that have not received the careful attention that they deserve. Two of the session had as their topics, issues we felt are primary forcing functions in the design of interactive systems capable of responding to, and responding in, natural language. These forcing functions involve the purpose of the interaction, the "social" conventions assumed by each participant, and the characteristics of the channel through which interaction takes place. The topic of the third session was the future of natural language communication with machines.

This workshop was held as a parasession in conjunction with the 18th Annual Meeting of the Association for Computational Linguistics (ACL). The three sessions were interleaved with the program for the ACL meeting. This allowed the participants of the ACL meeting attend these sessions and also permitted the invited participants of the special sessions to attend some or all of the program of ACL meeting.

A special committee was organized for the parasession. The members were:

Bonnie Lynn Webber, University of Pennsylvania, Organizer

Barbara Grosz, SRI

Jerry Hobbs, SRI.

The three sessions were as follows:

<u>Topic 1</u>:

Parasession Panel: Influence of the Problem Context

Barbara rosz, SRI, Chair

Wallace Chafe, University of California, Berkeley

Philip Cohen, BBN

Erving Goffman, University of Pennsylvania

Aravind K. Joshi, University of Pennsylvania

Charlotte Linde and J. A. Goguen, Structural Semantics and SRI

Deborah Tannen, Georgetown Unviersity

Topic 2:

Parasession Panel: Influence of the Social Context and Medium

Jerry Hobbs, SRI, Chair

John Carey, New York University

Phil Hayes, Carnegie Mellon University

Starr Roxanne Hiltz, Kenneth Johnson and Ann Marie Rabke, Upsala College

Emmanuel Schegloff, University of California at Los Angeles

John Thomas, IBM, T.I., Watson Research Center

Eleanor Wynn, Xerox Office Products Division

Topic 3:

Parasession Panel: Future Prospects

Bonnie Lynn Webber, University of Pennsylvania, Chair

Larry Harris, Artificial Intelligence Corporation

Gary Hendrix, SRI

Howard Morgan, University of Pennsylvania

A. Michael Noll, AT & T

Ben Shneiderman, University of Maryland

Murray Turoff, New Jersey Institute of Technology

The contributions to this workshop were published together with the proceedings of the ACL meeting. We have taken these contributions from the Proceedings of the ACL and assembled them together as a Proceedings of this workshop.

Defense Documentation Center Cameron Station Alexandrie, VA 22314	12 copies
Office of Naval Research Arlington, VA 22217	
Information System Program (437)	2 copies
Code 200	1 copy
Code 455	1 copy
Code 458	1 copy
Office of Naval Research Eastern/Central Regional Office Bldg 114, Section D 666 Summer Street Boston, MA 02210	1 copy
Office of Naval Research Brance Office, Chicago 536 South Clark Street Chicago, IL 60605	1 copy
Office of Naval Research Western Regional Office 1030 East Green Street Pasadena, CA 91106	1 copy
Naval Research Laboratory Technical Information Division, Cod3 2627 Washington, D.C. 20375	6 copies
Dr. A. L. Slafkosky Scientific Advisor Commandant of the Marine Corps (Code RD-1) Washington, D. C. 20380	1 copy
Naval Ocean Systems Center Advanced Software Technology Division Code 5200 San Diego, CA 92152	1 copy
Mr. E. H. Gleissner Naval Ship Research & Development Center Computation and Mathematics Department Bethesda, MD 20084	1 copy
Captain Grace M. Hopper (008) Naval Data Automation Command Washington Navy Yard Building 166 Washington, D. C. 20374	1 copy

Office of Naval Research Resident Representative University of Pennsylvania David Rittenhouse Laboratory 209 S. 33rd Street Philadelphia, Pa. 19104	1 copy
Dr. Marvin Denicoff Program Director Information Systems Department of the Navy Office of Naval Research Arlington, Virginia 22217	1 copy
Dr. Bonnie Webber Department of Computer and Information Science University of Pennsylvania The Moore School of Electrical Engineering D2 Philadelphia, Pa. 19104	2 copies
Dr. Aravind K. Joshi Professor and Chairman Department of Computer and Information Science The Moore School of Electrical Engineering D2 University of Pennsylvania Philadelphia, Pa. 19104	2 copies

Interactive Discourse: Influence of Problem Context
Panel Chair's Introduction

Barbara Grosz
SRI International

The purpose of the special parasession on "Interactive Man/Machine Discourse" is to discuss some critical issues in the design of (computer-based) interactive natural language processing systems. This panel will be addressing the question of how the purpose of the interaction, or "problem context" affects what is said and how it is interpreted. Each of the panel members brings a different orientation toward the study of language to this question. My hope is that looking at the question from these different perspectives will expose issues critical to the study of language in general, and to the construction of computer systems that can communicate with people in particular. Of course, the issue of the influence of "problem context" is separable from the issue of how one might get a computer system to take into account the effects of this context (and, yes, even whether that is possible). My hope is that those on the panel who are concerned with the construction of computer-based natural language processing systems will address some of the issues of "how" and that all of the panelists will consider the prior questions of what effects there are and what general principles underlie how the "problem context" influences a dialogue.

There are two separate aspects to the "problem context" that influence the participants' expectations and hence their utterances: (1) the function of the discourse, and, (2) the domain of discourse.

Function: This aspect of the problem context concerns why the speaker and hearer are communicating and their relative roles in the communication. Casual conversations, classroom discussions, task-oriented dialogues, and stories have very different functions. Although it is most reasonable to consider computer systems as participating in a restricted kind of dialogue (namely, a dialogue which arises from aiding a person in the solution of some problem), it is still clear that such systems may assume different roles, e.g., that of an expert (user is an apprentice), tutor (student), or supplier of information (e.g., from a large data base). Each of the different functions results in different kinds of goals (e.g., teaching requires a different kind of informing than simple question answering) and each of the different roles will create different expectations on the part of the user and different needs in terms of the kinds of information the system has about the user.

Domain: This aspect concerns what a speaker is talking about, the subject matter of the discourse. The structure of the information being discussed has an effect on the language (cf. Chafe's "The Flow of Language and the Flow of Thought", Linde's work on apartment descriptions and planning, my work on focusing in task-oriented dialogues).

Both of these aspects of "problem context" have global effects on what gets discussed and in what "units", and local effects on how speakers express the information they convey. Clearly the two aspects interact. For example, what a speaker chooses to discuss next depends both on why he is telling the hearer and on the information itself and what it is related to.

Some questions to consider:

In what ways are the effects of problem context manifest in individual utterances and larger discourse units? How do people's "conversational styles" differ?

The above discussion of "function" gave several examples. There is no taxonomy of function (as I've used the word). How might such a taxonomy be constructed and used?

What kinds of expectations are set up by different kinds of functions?

What assumptions about the knowledge, beliefs, and goals that are shared by the participants are made by the different functions?

How do the constraints from function interact with those of domain?

What kinds of "tools" are useful for examining such issues? (e.g., what kinds of analysis of data can be done)?

What happens when expectations generated by problem context (either function or domain) are violated?

SHOULD COMPUTERS WRITE SPOKEN LANGUAGE?

Wallace L. Chafe
University of California, Berkeley

Recently there has developed a great deal of interest in the differences between written and spoken language. I joined this trend a little more than a year ago, and have been exploring not only what the specific differences are, but also the reasons why they might exist. The approach I have taken has been to look for differences between the situations and processes involved in speaking on the one hand and writing on the other, and to speculate on how those differences might be responsible for the observable differences in the output. What happens when we write and what happens when we speak are different things, both psychologically and socially, and I have been trying to see how what we do in the two situations leads to the specific things that we find in writing and speaking.

I occasionally interact with the UNIX computer system at Berkeley, for various purposes. In the context of my concern about differences between writing and speaking, I have begun to wonder whether the kind of communication we are used to receiving from computers is more like writing or speaking. You may think that computers obviously write to us. They send us messages that we can read off of a cathode ray tube, or that get printed out for us on a piece of paper. In that respect what computers produce is written language. But it comes at us in a way that is very different from the way written language usually does. Usually we are faced with a printed page on which the writing is all there, and has been there for a long time. The temporal process by which the writing was put there has absolutely no relevance to us as we peruse the page at our leisure. The timing of our reading is in no way controlled by the timing by which the words were entered on the page. My computer terminal, on the other hand, is steadily chugging away, producing language before my eyes at the rate of 30 characters a second. Under some circumstances I could wait until it had produced a whole page before I began to read. But I don't usually do that. I eagerly follow the steady flow of letters as they appear, just as I would eagerly listen to the spoken sounds of someone who was telling me something I wanted to know. This processing in real time seems in that respect more like spoken language, although what is being produced is written. Furthermore, the computer system and I often, indeed characteristically, engage in quick exchanges, much like conversations, which is not what I am accustomed to doing with written language. So I want to suggest that when it is looked at from the point of view of the dichotomy between written and spoken language, the computer language we normally deal with is neither fish nor foul. It is produced in written form, but on the other hand it is produced in real time, and we are able to respond and interact as we are not able to do with a printed page.

Recent work seems to have shown that there are a number of features which are characteristic of spoken language, and a number of other features characteristic of written. It is not that spoken language never contains any of the features of writtenness, or that written language never contains any of the features of spokenness. It is only that certain features tend to be associated with one or the other medium, and that the features become more polarized as one approaches the extremes of colloquialness on the one hand, or of literariness on the other. In between one finds various mixtures of literary talk and conversational writing.

In looking for reasons why these distinguishing features exist, I have found it useful to attribute some of them to the temporal differences between writing and speaking, and some of them to the interactional differences.

Temporally, writing as an activity is much slower than speaking. Speaking seems to be produced one "idea unit" at a time, each idea unit having a mean length of about 2 seconds, or 6 words. Every so often a sequence of idea units ends in a falling pitch intonation of the sort we identify with the ending of a sentence. Pauses usually occur between idea units, and longer pauses between sentences. The idea units within a spoken sentence tend to be strung together in a coordinate fashion, typically with the word "and" appearing as a link. There is little of the fancy syntax we find in written language, by which some idea units are subordinated to and embedded within others. It has been hypothesized that speakers' attention capacities are not great enough to allow them to engage in much elaborate syntax. The flow of idea units is enough to keep them occupied. Writing, on the other hand, is peculiar in that the process of writing itself occupies an inordinate amount of time, even though, once we get past the first grade, it doesn't require a great deal of attention. Thus, writers have a lot of extra time and attention available to them, and apparently they often use it to construct elaborate sentences. As a result, whereas the sentences of spoken language have a distinctly fragmented quality, those of written language tend to be more integrated, with much more attention paid to subordinating idea units within others in complex ways. This integration vs. fragmentation dimension seems to be at the root of a number of the features which distinguish writing from speaking.

The other dimension I have been interested in seems to result from the different relation writers and speakers have to their respective audiences. Whereas speakers can interact directly with their listeners, obtaining ongoing confirmation, contradiction, and feedback, writers cannot normally do so, but are constrained to pay more attention to producing something that will stand on its own feet when it is read by someone later on in a different place. We can speak of the greater involvement of speakers, as contrasted with the greater detachment of writers. Many of the specific features distinguishing speaking and writing can be lined up on this involvement vs. detachment dimension.

How can a computer produce language that is maximally congenial to us humans, given the familiarity we already have with the characteristics of spoken and written language? What kind of human language should a computer simulate, in order that we can process it most easily? And to what extent is a computer able to produce such a simulation?

Let's play with the assumption that we human users would feel most at home with a computer terminal with which we could converse in something resembling human conversation, as close as this can be approximated by a machine which (1) can't yet make satisfactory sounds, but has to write what it says; and (2) doesn't know how to experience involvement with a human being. Let's consider what this machine would need to do to make us feel that we were interacting in something like the way we interact when we use spoken language.

Timing is one of the important factors. Instead of steadily producing letters at the rate of 30 a second, this machine might try producing language as spoken language is produced in real time. That would mean doing it at half the speed, for one thing: 15 characters a second would be about normal for the way we assimilate spoken language, and perhaps the rate at

which we naturally take in information But we would not want it spitting out one letter at a time at a steady rate, as it does now. That has little to do with the way we take in language, either spoken or written, under normal circumstances. Perhaps it should give us one word at a time, but I think it more likely that we would feel most comfortable with syllables: syllables timed to simulate the timing of syllables in normal English speech. Roughly speaking, stressed syllables would be longer and unstressed syllables shorter. A careful study of the timing of natural speech could introduce more sophistication here. At the end of each idea unit -- on the average after every 6 words -- there would be at least a brief pause, signaling the boundary of the idea unit and allowing time for processing. At the end of a sentence -- on the average after every 3 idea units -- the pause would be longer, and paragraph boundaries would be signaled by longer pauses. Idea units would be relatively fragented. Many of them would be connected by "and," and there would be little of the elaborate syntax one tends to find in written language.

As for involvement, the computer would need to learn that humans are imperfect recipients of information, and that redundancy and requests for confirmation are among the important devices to be used frequently in communicating with them. Frequent direct reference to the addressee is another feature of involvement that the computer could easily learn to use.

My terminal recently told me the following, at 30 steady characters per second:

> The "netlpr" command, when executed between computer center machines, now sets the ownership of net queue files correctly so that "netrm" will remove them and they are listed by the "netq" command.

While this is reasonably good written language, and comprehensible as such, I am asking whether meaningful linguistic interaction in real time might not better proceed something as follows, where you can imagine syllables being timed as they are timed in spoken English, brief pauses at the ends of lines, and longer pauses where I have double-spaced (T is the terminal and U the user):

T: Want to know about the "netlpr" command,
where you type in "netlpr"?

U: Sure.

T: You can just use it between computer center
machines,
OK?

Only if you're up here.

U: Yeah,
I know.

T: OK.

It'll show you who owns net queue files,
if you want to know that.

You can use "netrm" to get rid of them,
and you can get them listed with "netq".

That clear?

U: Yeah.

One problem with this is that the user has to type in at his or her normal typing rate, which will inevitably be much slower than speaking. But even so, the fragmentation and involvement which make this machine's output more like spoken language might significantly increase the user's comfort and comprehension. To know whether that is really true calls for further detailed research on the features which distinguish spoken from written language, and tests of whether the introduction of such features into computer language indeed makes a difference. Such research ought in any case to be rewarding beyond the bounds of this particular application.

Signalling the Interpretation of Indirect Speech Acts

Philip R. Cohen
Center for the Study of Reading
University of Illinois, &
Bolt, Beranek and Newman, Inc.
Cambridge, Mass.

This panel was asked to consider how various "problem contexts" (e.g., cooperatively assembling a pump, or Socratically teaching law) influence the use of language. As a starting point, I shall regard the problem context as establishing a set of expectations and assumptions about the shared beliefs, goals, and social roles of those participants. Just how people negotiate that they are in a given problem context and what they know about those contexts are interesting questions, but not ones I shall address here. Rather, I shall outline a theory of language use that is sensitive to those beliefs, goals, and expectations.

The theory is being applied to characterize actual dialogues occurring in the familiar task-oriented situation [10], in which an expert instructs a novice to do something, in our case to assemble a toy water pump. In such circumstances, the dialogue participants can be viewed as performing speech acts planned, primarily, to achieve goals set by the task. Other contexts undoubtedly emphasize the instrumental uses of language (e.g., [11]) but those problem contexts will not be considered here. The application of a model of speech act use to actual dialogue stresses the need for sources of evidence to substantiate predictions. The purpose of this paper is to point to one such source -- speaker-reference [9].

The natural candidate for a theory of instrumental use of speech acts is an account of rational action [12] -- what is typically termed "planning". However, contrary to the assumption of most planning systems, we are interested in the planning of (usually) cooperative agents who attempt to recognize and facilitate the plans of their partners* [1,4,5,16,20]. Such helpful behavior is independent of the use of language, but is the source of much conversational coherence.

A plan based theory of speech acts specifies that plan recognition is the basis for inferring the illocutionary force(s) of an utterance. The goal of such a theory is to formalize the set of possible plans underlying the use of particular speech acts to achieve a given set of goals. In light of the independent motivation for plan generation and recognition, such a formalism should treat communicative and non-communicative acts uniformly, by stating the communicative nature of an illocutionary act as part of that act's definition. A reasoning system, be it human or computer, would then not have to employ special knowledge about communicative acts; it would simply attempt to achieve or recognize goals.

The components of speech act planning and recognition systems developed so far include: a formal language for describing mental states and states of the physical and social worlds, operators for describing changes of state, associations of utterance features (e.g., mood) with certain operators, and a set of plan construction and recognition inferences. Illocutionary acts are defined as operators that primarily affect the mental states of speakers and hearers [3,8,13,17].

To be more specific, in the most fully developed attempt at such a theory, Perrault and Allen [13] show how plan recognition can "reason out" a class of indirect speech acts. Briefly, they define "surface" speech act operators, which depend on an utterance's mood, and operators for illocutionary acts such as requesting. Plan recognition involves inferences of the form "the agent intended to perform action X because he intended to achieve its effect in order to enable him to do some other action Y". Such inferences are applied to surface speech act operators (characterizing, for instance, "Is the salt near you?") to yield illocutionary operators such as requests to pass the salt.

The remainder of this paper attempts to illustrate the kinds of predictions made by the theory, and the use of anaphora to support one such prediction.* Consider the following dialogue fragment (transmitted over teletype) in the water pump context described earlier:

Expert: 1). "We need a clear bent tube for the bottom hole."
Novice: 2). "OK, it's done."
Expert: 3). "OK, now, start pumping"

The example is constructed to illustrate my point, but it does not "feel" artificial. Experiments we are conducting show analogous phenomena in telephone and teletype modes.

The theory predicts two inference paths for utterance 1 -- "helpful" and "intended". In the former case, the novice observes the surface-inform speech act indicated by a declarative utterance, and interprets it simply as an inform act that communicates a joint need. Then, because the novice is helpful, she continues to recognize the plan behind the expert's utterance and attempts to further it by performing the action of putting the spout over the hole. The novice, therefore, is acting on her own, evaluating the reasonableness of the plan inferred for the expert using private beliefs about the expert's beliefs and intentions. Alternatively, she could infer that the expert intended for it to be mutually believed that he intended her to put on the tube. Thus, the novice would be acting because she thinks the expert intended for her to do so. Later, she could summarize the expert's utterance and intentions as a request [7]. Perrault and Allen supply heuristics that would predict the preferred inference route to be the "intended" path since it is mutually believed that putting the tube on is the relevant act, and his intending that she perform pump-related acts is an expected goal in this problem context. To use Perrault and Allen's model for analyzing conversation, such predictions must be validated against evidence of the novice's interpretation of the expert's intent.

Signalling Interpretation of Intent

For this problem context and communication modality, the novice and expert shared knowledge that the expert will attempt to get the novice to achieve each subgoal of the physical task, and the novice must indicate successful completion of those subtasks. However, not all communicative acts achieving the goal of indicating successful completion provide evidence of the novice's interpretation of intent. For instance, the novice might say "I've put the bent tube on" simply to keep the expert informed of the situation. Such an informative act could arise if the problem context and prior conversation did not make the salience of putting the tube on mutually known. To supply evidence of the novice's interpretation of intent, her response must pragmatically presuppose that interpretation.

In our example, the novice has used "it" to refer to the action she has performed. It has been proposed that definite and pronominal/pro-verbal reference requires mutual belief that the object in question is in focus [10,15] and satisfies the "description" [6,14]. Assuming that the inferring of mutually believed goals places them in focus [15], the shared knowledge needed to refer using "it" is supplied by only one of the above interpretations -- the one summarizable as an indirect request.

* For this brief paper, I shall have to curtail discussion of the planning/plan recognition literature.

* Robinson [15] has identified this problem of reference to actions and has implemented a system to resolve them. In this paper, I stress the importance of that work to theories of speech act use.

Other signals of the interpretation of intent need to be identified to explain how the expert's "OK, now start pumping" communicates that he thinks she has interpreted him correctly -- mutual signalling of intent and its interpretation is central to conversational success.

A formal theory that could capture the belief, intention, and focus conditions for speaker-reference is thus clearly needed to validate models of speech act use. A plan-based theory might accomodate such an analysis via a decomposition of currently primitive surface speech acts to include reference acts [2,18]. By planning reference acts to facilitate the hearers' plans (cf. [4]), a system could perhaps also answer questions cooperatively without resorting to Gricean maxims or "room theories" [19].

I have given a bare bones outline of how a description of speaker-reference can serve as a source of empirical support to a theory of speech acts. However, much more research must take place to flesh out the theoretical connections. I have also deliberately avoided problems of computation here, but hope the panel will discuss these issues, especially the utility of computational models to ethnographers of conversation.

Acknowledgements:
I would like to thank Chip Bruce, Scott Fertig, and Sharon Oviatt for comments on an earlier draft.

References:

1. Allen, J. A plan-based approach to speech act recognition (Tech. Rep. No. 131/79). Toronto: University of Toronto, Department of Computer Science, January 1979.
2. Appelt, D. Problem-solving applied to language generation. (This volume).
3. Bruce, B. Belief systems and language understanding (BBN Report No. 2973). Cambridge, Mass.: Bolt, Beranek and Newman, January 1975.
4. Bruce, B., & Newman, D. Interacting plans. Cognitive Science, 1978, 2, 195-233.
5. Carbonell, J. G. Jr. POLITICS: Automated ideological reasoning. Cognitive Science, 1978, 2, 27-51.
6. Clark, H. H., & Marshall, C. Definite reference and mutual knowledge. In A. K. Joshi, I. A. Sag, & B. L. Webber (Eds.), Proceedings of the Workshop on Computational Aspects of Linguistic Structure and Discourse Setting. New York: Cambridge University Press, in press.
7. Cohen, P. R., & Levesque, H. L. Speech acts and the recognition of shared plans. In Proceedings: Annual meeting of the Canadian Society for the Computational Study of Intelligence, Victoria, B.C., 1980.
8. Cohen, P. R., & Perrault, C. R. Elements of a plan-based theory of speech acts. Cognitive Science, 1979, 3, 177-212.
9. Donnellan, K. Speaker references, descriptions, and anaphora. In P. Cole (Ed.), Syntax and semantics (Vol. 9): Pragmatics. New York: Academic Press, 1978.
10. Grosz, B. The representation and use of focus in dialogue understanding (Technical Note 151). Menlo Park, Calif.: Stanford Research Institute, Artificial Intelligence Center, July 1977.
11. Hobbs, J. R., & Evans, D. E. Conversation as planned behavior (Technical Note 203). Menlo Park, Calif.: Stanford Research Institute, Artificial Intelligence Center, 1979.
12. Morgan, J. L. Toward a rational model of discourse comprehension. In D. Waltz (Ed.), Proceedings: Theoretical Issues in Natural Language Understanding. Urbana: University of Illinois, Coordinated Science Laboratory, 1978.
13. Perrault, C. R., & Allen, J. F. A plan-based analysis of indirect speech acts. In submission.
14. Perrault, C. R., & Cohen, P. R. Inaccurate reference. In A. K. Joshi, I. A. Sag, & B. L. Webber (Eds.), Proceedings of the Workshop on Computational Aspects of Linguistic Structure and Discourse Setting. New York: Cambridge University Press, in press.
15. Robinson, A. E. The interpretation of verb phrases in dialogs (Technical Note 206). Menlo Park, Calif.: Stanford Research Institute, Artificial Intelligence Center, 1980.
16. Schank, R., & Abelson, R. Scripts, plans, goals, and understanding. Hillsdale, N.J.: Erlbaum, 1977.
17. Schmidt, C. F. Understanding human action. In Proceedings of the conference on Theoretical Issues in Natural Language Processing. Cambridge, Mass., 1975.
18. Searle, J. R. Speech acts: An essay in the philosophy of language. Cambridge: Cambridge University Press, 1969.
19. Shannon, B. Where-questions. In Proceedings of the Seventeenth Annual Meeting of the ACL, San Diego, 1979. Pp. 73-75.
20. Wilensky, R. Understanding goal-based stories (Research Rep. No. 140). New Haven, Conn.: Yale University, Department of Computer Science, September 1978.

PARASESSION ON TOPICS IN INTERACTIVE DISCOURSE

INFLUENCE OF THE PROBLEM CONTEXT*

Aravind K. Joshi

Department of Computer and Information Science
Room 268 Moore School
University of Pennsylvania
Philadelphia, PA 19104

My comments are organized within the framework suggested by the Panel Chair, Barbara Grosz, which I find very appropriate. All of my comments pertain to the various issues raised by her; however, wherever possible I will discuss these issues more in the context of the "information seeking" interaction and the data base domain.

The primary question is how the purpose of the interaction or "the problem context" affects what is said and how it is interpreted. The two separate aspects of this question that must be considered are the function and the domain of the discourse.

1. Types of interactions (functions):

1.1 We are concerned here about a computer system participating in a restricted kind of dialogue with a person. A partial classification of some existing interactive systems, as suggested by Grosz, is as follows. I have renamed the third type in a somewhat more general fashion.

	Participant P1 (Computer system)	Participant P2 (Person)
Type A	Expert	Apprentice
Type B	Tutor	Student
Type C	Information provider (some sort of large and complex data base or knowledge base)	Information seeker

Each type subsumes a variety of subtypes. For example, in type C, subtypes arise depending on the kind of information available and the type of the user. (More on this later when we discuss the interaction of constraints on function and domain).

1.2 It should be noted also that these different types are not really completely independent; information seeking (Type C) is often done by the apprentice (Type A) and student (Type B), and some of the explaining done by tutors (Type B) is also involved in the Type C interaction, for example, when P1 is trying to explain to P2 the structure of the data base.

1.3 The roles of the two participants are also not fixed completely. In the type C interaction, sometimes P2 partly plays the role of an expert (or at least appears to do so) believing that his/her expert advice may help the system answer the question more 'easily' or 'efficiently'. For example[1], in a pollution data base P1 may ask: Has company A dumped any wastes last week? and follow up with advice: Try arsenic first. In the expert-apprentice interaction, the expert's advice is assumed to be useful by the apprentice. In the data base domain it is not clear whether the 'expert' advice provided by the user is always useful. It does however provide information about the user which can be helpful in presenting the response in an appropriate manner; for example, if arsenic indeed was one of the wastes dumped, then, perhaps, it should be listed first.

1.4 The interactions of the type we are concerned about here are all meant to aid a person in some fashion. Hence, a general characterization of all these types is a helping function. However, it is useful to distinguish the types depending on whether an information seeking or information sharing interaction is involved. Type C interaction is primarily information seeking, although some sharing interaction is involved also. This is so because information sharing facilitates information seeking, for example[2], when P1 explains the structure of the data base to P2, so that P2 can engage in information seeking more effectively. Type A and B are more information sharing than information seeking interactions.

1.5 Another useful distinction is that type C interaction has more of a service function than types A and B which have more of a training function. Training involves more of information sharing, while service involves more of providing information requested by the user.

2. Information about the user:

2.1 By user we usually mean user type and not a specific user. User information is essential in determining expectations on the part of the user and the needs of the user. Within each type of interaction there can be many user types and the same information may be needed by these different types of users for different reasons. For example, in type C interaction, preregistration information about a course scheduled for the forthcoming term may be of interest to an instructor because he/she wants to find out how popular his/her course is. On the other hand, the same data is useful to the registrar for deciding on a suitable room assignment. The data base system will often provide different views of the same data to different user types.

2.2 In general, knowledge about the user is necessary, at least in the type C interaction in order to decide

(i) how to present the requested information,

(ii) what additional information, beyond that explicitly requested, might be usefully presented (this aspect is not independent of (i) above),

(iii) what kind of responses the system should provide when the user's misconceptions about the domain

* This work was partially supported by the NSF grant MCS79-08401.

I want to thank Eric Mays, Kathy McKeown, and Bonnie Webber for their valuable comments on an earlier draft of this paper.

(i.e., both the structure and content of the data base, in short, what can be talked about) are detected.

(More about this in section 5).

3. Conversational style:

3.1 In the type C interaction, the user utterances (more precisely, user's typewritten input) are a series of questions separated by the system's responses. By and large, the system responds to the current question. However, knowledge about the preceding interaction i.e., discourse context (besides, of course, the information about the user) is essential for tracking the "topic" and thereby determining the "focus" in the current question. This is especially important for determining how to present the answer as well as how to provide appropriate responses, when user's misconceptions are detected.

Type A and B interactions perhaps involve a much more structured dialogue where the structure has its scope over much wider stretches of discourse as compared to the dialogues in the type C interactions, which appear to be less structured.

3.2 The type of interaction involved certainly affects the conversational style; however, little is known about conversational style in interactive man/machine communication. Folklore has it that users adapt very rapidly to the system's capabilities. It might be useful to compare this situation to that of a person talking to a foreigner. It has been claimed that natives talking to foreigners deliberately change their conversational style[3] (for example, slowing down their speech, using single words, repeating certain words, and even occasionally adopting some of the foreigner's style, etc.). It may be that users treat the computer system as an expert with respect to the knowledge of the domain but lacking in some communicative skills, much like a native talking to a foreigner.

Perhaps it is misleading to treat man/machine interactive discourse as just (hopefully better and better) approximations to human conversational interactions. No matter how sophisticated these systems become, they will at the very least lack the face to face interaction. It may be that there are certain aspects of these interactions that are peculiar to this modality and will always remain so. We seem to know so little about these aspects. These remarks, perhaps, belong more to the scope of the panel on social context than to the scope of this panel on the problem context.

4. Relation of expectations and functions:

4.1 In the information seeking interaction, usually, the imperative force of the user's questions is to have the system bring it about that the user comes to know whatever he/she is asking for. Thus in asking the question Who is registered in CIS 591? the user is interested in knowing who is registered in CIS 591. The user is normally not interested in how the system got the answer. In the type A and B interactions the imperative force of a question from the user (apprentice or student) can either be the same as before or it can have the imperative force of making the system show the user how the answer was obtained by the system.

4.2 In the data base domain, although, primarily the user is interested in what the answer is and not in how it was obtained, this need not be the case always. Sometimes the user would like to have the answer accompanied by how it was obtained, the 'access paths' through the data base, for example.

4.3 Even when only the what answer is expected, often the presentation of the answer has to be accompanied by some 'supportive' information to make the response useful to the user[4]. For example, along with the student name, his/her department or whether he/she is a graduate or undergraduate student would have to be stated. If telephone numbers of students are requested then along with the telephone numbers, the corresponding names of students will have to be provided.

5. Shared knowledge and beliefs:

5.1 The shared beliefs and goals are embodied in the system's knowledge of the user (i.e., a user model). It is important to assume that not only the system has the knowledge of the user but that the user assumes that the system has this knowledge. This is very necessary to generate appropriate cooperative responses and their being correctly understood as such by the user. In ordinary conversations this type of knowledge could lead to an infinite regress and hence, the need to require the shared knowledge to be 'mutual knowledge'. However, in the current data base systems (and even in the expert-apprentice and tutor-student interactions) I am not aware of situations that truly lead to some of the well known problems about 'mutual knowledge'.

5.2 As regards the knowledge of the data base itself (both structure and content), the system, of course, has this knowledge. However, it is not necessary that the user has this knowledge. In fact very often the user's view of the data base will be different from the system's view. For large and complex data bases this is more likely to be the case. The system has to be able to discern the user's view and present the answers, keeping in mind the user's view, while insuring that his/her view is consistent with the system's view.

5.3 When the system recognizes some disparity between its view and the user's view, it has to provide appropriate corrective responses. Users' misconceptions could be either extensional (i.e., about the content of the data base) or intensional (i.e., about the structure of the data base)[4]. Note that the extensional/intensional distinction is from the point of view of the system. The user may not have made the distinction in that way. Some simple examples of corrective responses are as follows. A user's question: Who took CIS 591 in Fall 1979? presumes that CIS 591 was offered in Fall 1979. If this was not the case then a response None by the system would be misleading; rather the response should be that CIS 591 was not offered in Fall 1979. This is an instance of an extensional failure. An example of intensional failure is as follows. A user's question: How many undergraduates taught courses in Fall 1979? presumes (among other things) that undergraduates do teach courses. This is an intensional presumption. If it is false then once again an answer None would be misleading; rather the response should be that undergraduates are not permitted to teach courses, faculty members teach courses, and graduate students teach courses. The exact nature of this response depends on the structure of the data base.

6. Complexity of the domain:

6.1 In each type of interaction the complexity of the interaction depends both on the nature of the interaction (i.e., function) as well as the domain. In many ways the complexity of the interaction ultimately seems to depend on the complexity of the domain. If the task itself is not very complex (for example, boiling water for tea instead of assembling a pump) the task oriented expert-apprentice interaction cannot be very complex. On the other hand data base interaction which appear to be simple at first sight become in-

creasingly complex when we begin to consider (i) dynamic data bases (i.e., they can be updated) and the associated problems of monitoring events (ii) data bases with multiple views of data, (iii) questions whose answers require the system to make fairly deep inferences and involve computations on the data base i.e., the answers are not obtained by a straightforward retrieval process, etc.

NOTES:

1. As in the PLIDIS system described by Genevieve Berry-Rogghe.

2. As in Kathy McKeown's current work on generating descriptions and explanations about data base structure.

3. For example, by R. Rammurti in her talk on 'Strategies involved in talking to a foreigner' at the Penn Linguistics Forum 1980 (published in Penn Review of Linguistics, Vol. 4, 1980).

4. Many of my comments about supportive information and corrective responses when misconceptions about the content and the structure of the data base are detected are based on the work of Jerry Kaplan and Eric Mays.

ON THE INDEPENDENCE OF DISCOURSE STRUCTURE AND SEMANTIC DOMAIN

Charlotte Linde* J.A. Goguen+ *

1. THE STATUS OF DISCOURSE STRUCTURE

Traditionally, linguistics has been concerned with units at the level of the sentence or below, but recently, a body of research has emerged which demonstrates the existence and organization of linguistic units larger than the sentence. (Chafe, 1974; Goguen, Linde, and Weiner, to appear; Grosz, 1977; Halliday and Hasan, 1976; Labov, 1972; Linde, 1974, 1979, 1980a, 1980b; Linde and Goguen, 1978; Linde and Labov, 1975; Polanyi, 1978; Weiner, 1979.) Each such study raises a question about whether the structure discovered is a property of the organization of language or whether it is entirely a property of the semantic domain. That is, are we discovering general facts about the structure of language at a level beyond the sentence, or are we discovering particular facts about apartment layouts, water pump repair, Watergate politics, etc? Such a crude question does not arise with regard to sentences. Although much of the last twenty years of research in sentential syntax and semantics has been devoted to the investigation of the degree to which syntactic structure can be described independently of semantics, to our knowledge, no one has attempted to argue that all observable regularities of sentential structure are attributable to the structure of the real world plus general cognitive abilities. Yet this claim is often made about regularities of linguistic structure at the discourse level. In order to demonstrate that at least some of the structure found at the discourse level is independent of the structure of the semantic domain, we may show that there are discourse regularities across semantic domains. As primary data, we will use apartment layout description, small group planning, and explanation. These have all been found to be discourse units, that is, bounded linguistic units one level higher than the sentential level, and have all been described within the same formal theory. It should be noted that we do not claim that the structures found in these discourse units is entirely independent of structure of the semantic domain, because of course the structure of the domain has some effect.

2. TREE TRANSFORMATIONS IN DISCOURSE PRODUCTION

The discourse units mentioned above have all been found to be tree structured. This is a claim that any such discourse can be divided into parts such that there are significant relations of dominance among these parts. These trees can be viewed as being constructed by a sequence of transformations on an initial empty tree, with each transformation corresponding to an utterance by participants, which may add, delete, or move nodes of the tree. The sequence of transformations encodes the construction of the discourse as it actually proceeds in time.

We now turn to a discussion of the discourse units which have been analysed according to this model.

* Structural Semantics, P.O. Box 707, Palo Alto, California 94302.

+ SRI International, 333 Ravenswood Ave., Menlo Park, California 94025.

2.1 SPATIAL DESCRIPTIONS AS TOURS

In an investigation of the description of spatial networks, speakers were asked to describe the layout of their apartment. The vast majority of speakers used a "tour strategy," which takes the hearer on an imaginary tour of the apartment, building up the description of the layout by successive mention of each room and its position. This tour forms a tree composed of the entry to the apartment as root with the rooms and their locations as nodes, and with an associated pointer indicating the current focus of attention, expressed by unstressed <u>you</u>.

It might be argued that the tree structure of these descriptions is a consequence of the structure of apartments rather than of the structure of discourse. However, there are apartments which are not tree structured, because some rooms have more than one entrance, thus allowing multiple routes to the same point; but in their descriptions, speakers traverse only one route; that is, loops in the apartment are always cut in the descriptions.[1] Thus, although some of the tree structure may be attributable to the physical structure being described, some of it is a consequence of the ease of expressing tree structures in language, and the difficulty of expressing graph structures.

The tree structure of apartment descriptions is constructed using only addition transformations, and pointer movement transformations (called "pops" in Linde and Goguen (1978)) which bring the focus of attention back from a branch which has been traversed to the point of branching. The construction of the tree is entirely depth first.

2.2 SPATIAL DESCRIPTIONS AS MAPS

In describing apartment layouts, there is a minority strategy, used by 4% of the speakers (3 out of 72 cases of the data of Linde (1974)) describing the layout in the form of a map. The speaker first describes the outside shape, then sketches the internal spatial divisions, and finally labels each internal division. This strategy can also be described as a tree construction, in this case, a breadth first traversal with the root being the outside shape, the internal divisions the next layer of nodes, and the names of these divisions the terminal nodes. Because there are so few example, it is not possible to give a detailed description of the rules for construction.

2.3 PLANNING

We have argued that the structure of apartment layout descriptions is not entirely due to the structure of the semantic domain; however, a question remains as to whether it is the restriction to a limited domain which permits precise description. To investigate this, let us consider the Watergate transcripts, which offer a spectacularly unrestricted semantic domain, specifically those portions in which the president and his advisors engage in the activity of planning. (Linde and Goguen, 1978). Planning sessions form a discourse unit with

[1] In more mathematical language, the linear sequence of rooms is the depth first traversal of a minimal spanning tree of the apartment graph.

discernable boundaries and a very precisely describable internal structure. Although we can not furnish any detailed description of the semantic domain, we can be extremely precise about the social activity of plan construction.

Because the cases we have examined involve planning by a small group, the tree is not constructed exclusively by addition, as are the types discussed above. Deletion, substitution, and movement also occur, as a plan is criticised and altered by all members of the group.

2.4 EXPLANATION

A discourse unit similar to planning is explanation. (Weiner, 1979; Goguen, Linde and Weiner to appear.) (By explanation we here include only the discourse unit of the form described below; we exclude discourse units such as narratives or question-response pairs which may socially serve the function of explanation.) Informally, explanation is that discourse unit which consists of a proposition to be demonstrated, and a structure of reasons, often multiply embedded reasons, which support it. The data of this study are accounts given of the choice to use the long or short income tax form, explanations of career choices, and material from the Watergate transcripts in which an evaluation is given of how likely a plan is to succeed, with complex reasons for this evaluation.

Like apartment descriptions and small group planning, explanation can be described as the transformational construction of a tree structure. Since in the casesexamined, a single person builds the explanation, there are no reconstructive transformations such as deletion or movement of subtrees; the transformations found are addition and pointer movement. Pointer movement is particularly complex in this discourse unit since explanation permits embedded alternate worlds, which require multiple pointers to be maintained. Explanation structure appears to be the same in the three different semantic domains, suggesting that the discourse structure is due to genral rules plus a particular social context, rather than being due to the structure of the semantic domain.

3. CRITERIA FOR EVALUATING DISCOURSE STRUCTURES

The criticism might be made of these tree structures that an analyst can impose a tree structure on any discourse, without any proof that it is related to what the speaker himself was doing. We would claim that although we have, of course, no direct access to the cognitive processes of speakers, there are two related criteria for evaluating a proposed discourse structure.

3.1 TEXT MARKING

One criterion for judging the relative naturalness of a particular analysis is the degree to which the text being analysed contains markers of the structure being postulated. Thus, we have some confidence that the speaker himself is proceeding in terms of a branching structure when we find markers like "Now as you're coming into the front of the apartment, if you go straight rather than go right or left, you come into a large living room area," or "On the one hand, we could try ..." The opposite case would be a text in which the divisions postulated by an analyst on the basis of some a priori theory had no semantic or syntactic marking in the text.

3.2 FRUITFULNESS OF THE ANALYSIS

A second criterion is whether some postulated structure is fruitful in generating further suggestions for how to explore the text. Thus, the tree analyses of apartment layout descriptions, planning, and explanation, give rise to questions such as how various physical layouts are turned into trees, how trees are traversed, the social consequences of particular transformations, the apparent psychological ease or difficulty of various transformations, the relation of discourse structure to syntactic structure, etc. (see Linde and Goguen, 1978) By contrast, an unfruitful analysis will give rise to few or no interesting research questions, and will not permit the analyst to investigate questions about the discourse unit which he or she has reason to believe are interesting.

4. GENERAL PRINCIPLES OF DISCOURSE STRUCTURE

Given that these postulated structures are useful models of what speakers do, we may ask how it is that speakers produce texts with these structures. It is known that children must learn to produce well-formed narratives. It might be hypothesized that each discourse unit must be separately learned, and that each has its own unrelated set of rules. However, there is evidence that there are very general rules for discourse construction, which hold across discourse units, and which can be used to construct novel discourse units. The test case for such a hynothesis is the production of a discourse unit which is not a part of speakers' ordinary repetoire, but rather, is made up for the occasion of the experiment. Such an experiment was performed by asking people to describe the process of getting themselves and their husbands and children off to work in the morning. (Linde, in preparation) These "morning routines" are typically well-structured and regular; everyone appears to do them the same way. We know that the speakers had never produced such discourses before, since we never in ordinary discourse hear such extended discussions of the details of daily life. (Even bores have their limits.) Therefore, the regularities must be the product of the intersection of a particular real world domain, in this case, multiple parallel activities, with very general rules for discourse construction.[2]

4.1 META-RULES OF DISCOURSE STRUCTURE

We are by no means ready to offer a single general theory of discourse structure; that must wait until a sufficiently large number of discourse types has been investigated in detail. However, the following rules have been observed in two or more discourse units, and it is rules of this type that we would like to investigate in other discourse units.

1. The most frequent subordinator for a given discourse unit will have the most minimal marking in the text, most frequently being marked with lexical <u>and</u>. Moreover, it will not be necessary to establish this node before beginning the first branch, but only when the return to the branch point is effected.

2. All other node types which subordinate two or more branches, such as <u>exclusive or</u> or <u>conditional</u>, must be indicated by markers in the text before the first branch is begun.

3. Depth-first traversal is the most usual strategy.

4. Pop markers are available to indicate return to a branch point or higher node; it is never necessary to recapitulate in reverse the entire traversal of a branch.

---

[2] This is interesting for the light which it sheds on natural structures for the description of concurrent activities.

5. CONCLUSIONS

The reason for being interested in regularities of discourse structure, particularly regularities which hold across a number of discourse types, is that they suggest universals of what is often called "mind," and, more practically, they also suggest features which might be part of systems for language understanding and production. Indeed Weiner (to appear) has constructed a system for the production of explanations of U.S. income tax law based on the transformational theory of explanation discussed in section 2.4. There is, moreover, the possibility of designing meta-systems, which might be programmed to handle a variety of discourse types.

ACKNOWLEDGEMENTS

We would like to thank R. M. Burstall and James Weiner for their help throughout much of the work reported in this paper. We owe our approach to discourse analysis to the work of William Labov, and our basic orientation to Chogyam Trungpa, Rinpoche.

REFERENCES

Chafe, Wallace, 1974. Language and Consciousness. Language, Vol. 50, 111-133.

Goguen, J.A., Charlotte Linde, and James Weiner. to appear. The Structure of Natural Explanation.

Grosz, Barbara J. 1977. The Representation and Use of Focus in Dialogue Understanding. SRI Technical Note 151.

Halliday, M.A.K. and Ruqaiya N. Hasan, 1976. Cohesion in English, Longman, London.

Labov, William, 1972. The Transformation of Experience into Narrative Syntax, in Language in the Inner City, Philadelphia, University of Pennsylvania Press.

Linde, Charlotte, 1974. The Linguistic Encoding of Spatial Information. Columbia University, Department of Linguistics dissertation.

Linde, Charlotte, 1979. Focus of Attention and the Choice of Pronouns in Discourse, in Syntax and Semantics, Vol.12 Discourse and Syntax, ed. Talmy Givon, Academic Press, New York.

Linde, Charlotte, 1980a. The Organization of Discourse, in The English Language in its Social and Historical Context ed. Timothy Shopen, Ann Zwicky and Peg Griffen, Winthrop Press, Cambridge, Massachusetts.

Linde Charlotte, 1980b. The Life Story: A Temporally Discontinuous Discourse Type, in Papers From the Kassel Workshop on Psycholinguistic Models of Production.

Linde, Charlotte, in preparation. The Discourse Structure of the Description of Concurrent Activity.

Linde, Charlotte and J.A. Goguen, 1978. The Structure of Planning Discourse, Journal of Social and Biological Structures, Vol. 1, 219-251.

Linde, Charlotte and William Labov, 1975. Spatial Networks as a Site for the Study of Language and Thought, Language, Vol. 51, 924-939.

Polanyi, Livia, 1978. The American Story. University of Michigan Department of Linguistics dissertation.

Weiner, James, 1979. The Structure of Natural Explanation: Theory and Application. System Development Corporation, SP-4035.

Weiner, J. BLAH: A System Which Explains its Reasoning, to appear in Artificial Intelligence.

The Parameters of Conversational Style

Deborah Tannen
Georgetown University

There are several dimensions along which verbalization responds to context, resulting in individual and social differences in conversational style. Style, as I use the term, is not something extra added on, like decoration. Anything that is said must be said in some way; co-occurrence expectations of that "way" constitute style. The dimensions of style I will discuss are:

1. Fixity vs. novelty
2. Cohesiveness vs. expressiveness
3. Focus on content vs. interpersonal involvement.

Fixity vs. novelty

Any utterance or sequence must be identified (rightly or wrongly, in terms of interlocutor's intentions) with a recognizable frame, as it conforms more or less to a familiar pattern. Every utterance and interaction is formulaic, or conventionalized, to some degree. There is a continuum of formulaicness from utterly fixed strings of words (situational formulas: "Happy birthday," "Welcome home," "Gezundheit") and strings of events (rituals), to new ideas and acts put together in a new way. Of course, the latter does not exist except as an idealization Even the most novel utterance is to some extent formulaic, as it must use familiar words (witness the absurdity of Humpty Dumpty's assertion that when he uses a word it means whatever he wants it to mean, and notice that he chooses to exercise this license with only one word); syntax (again Lewis Carroll is instructive: the "comprehensibility" of Jabberwocky); intonation; coherence principles (cf Alton Becker); and content (Mills' "vocabularies of motives," e.g.). All these are limited by social convention. Familiarity with the patterns is necessary for the signalling of meaning both as prescribed and agreed upon, and as cued by departure from the pattern (cf Hymes).

For example, a situational formula is a handy way to signal familiar meaning, but if the formula is not known the meaning may be lost entirely, as when a Greek says to an American cook, "Health to your hands." If meaning is not entirely lost, at least a level of resonance is lost, when reference is implicit to a fixed pattern which is unfamiliar to the interlocutor. For example, when living in Greece and discussing the merits of buying an icebox with a Greek friend, I asked, "Doesn't the iceman cometh?" After giggling alone in the face of his puzzled look, I ended up feeling I hadn't communicated at all. Indeed I hadn't.

Cohesiveness vs. expressiveness

This is the basic linguistic concept of markedness and is in a sense another facet of the above distinction. What is prescribed by the pattern for a given context, and what is furnished by the speaker for this instance? To what extent is language being used to signal "business as usual," as opposed to signalling, "Hey, look at this!" This distinction shows up on every level of verbalization too: lexical choice, pitch and amplitude, prosody, content, genre, and so on. For example, if someone uses an expletive, is this a sign of intense anger or is it her/his usual way of talking? If they reveal a personal experience or feeling, is that evidence that you are a special friend, or do they talk that way to everybody? Is overlap a way of trying to take the floor away from you or is it their way of showing interest in what you're saying? Of course, ways of signalling special meaning -- expressiveness -- are also prescribed by cultural convention, as the work of John Gumperz shows. The need to distinguish between individual and social differences is thus intertwined with the need to distinguish between cohesive and expressive intentions. One more example will be presented, based on spontaneous conversation taped during Thanksgiving dinner, among native speakers of English from different ethnic and geographic backgrounds.

In responding to stories and comments told by speakers from Los Angeles of Anglican/Irish background, speakers of New York Jewish background often uttered paralinguistically gross sounds and phrases ("WHAT!?" "How INTeresting!" "You're KIDding!" "Ewwwwww!"). In this context, these "exaggerated" responses had the effect of stopping conversational flow. In contrast, when similar responses were uttered while listening to stories and comments by speakers of similar background, they had the effect of greasing the conversational wheels, encouraging conversation. Based on the rhythm and content of the speakers' talk, as well as their discussion during playback (i.e. listening to the tape afterwards), I could hypothesize that for the New Yorkers such "expressive" responses are considered business as usual; an enthusiasm constraint is operating, whereby a certain amount of expressiveness is expected to show interest. It is a cohesive device, a conventionally accepted way of having conversation. In contrast, such responses were unexpected to the Californians and therefore were taken by them to signal, "Hold it! There's something wrong here." Consequently, they stopped and waited to find out what was wrong. Of course such differences have interesting implications for the ongoing interaction, but what is at issue here is the contrast between the cohesive and expressive use of the feature.

Focus on content vs. interpersonal involvement

Any utterance is at the same time a statement of content (Bateson's 'message') and a statement about the relationship between interlocutors ('metamessage'). In other words, there is what I am saying, but also what it means that I am saying this in this way to this person at this time. In interaction, talk can recognize, more or less explicitly and more or less emphatically (these are different), the involvement between interlocutors. It has been suggested that the notion that meaning can stand alone, that only content is going on, is associated with literacy, with printed text. But certainly relative focus on content or on interpersonal involvement can be found in either written or spoken form. I suspect, for example, that one of the reasons many people find interaction at scholarly conferences difficult and stressful is the conventional recognition of only the content level, whereas in fact there is a lot of involvement among people and between the people and the content. Whereas the asking of a question following a paper is conventionally a matter of exchange of information, in fact it is also a matter of presentation of self, as Goffman has demonstrated for all forms of behavior.

A reverse phenomenon has been articulated by Gail Dreyfuss. The reason many people feel uncomfortable, if not scornful, about encounter group talk and "psychobabble" is that it makes explicit information about relationships which people are used to signalling on the meta level.

Relative focus on content gives rise to what Kay (1977) calls "autonomous" language, wherein maximal meaning is encoded lexically, as opposed to signalling it through use of paralinguistic and nonlinguistic channels, and wherein maximal background information is furnished, as opposed to assuming it is already known as a consequence of shared experience. Of course this is an idealization as well, as no meaning at all could be communicated if

there were no common experience, as Fillmore (1979) amply demonstrates. It is crucial, then, to know the operative conventions. As much of my own early work shows, a hint (i.e. indirect communication) can be missed if a listener is unaware that the speaker defines the context as one in which hints are appropriate. What is intended as relatively direct communication can be taken to mean far more, or simply other, than what is meant if the listener is unaware that the speaker defines the context as one in which hints are inappropriate. A common example seems to be communication between intimates in which one partner, typically the female, assumes, "We know each other so well that you will know what I mean without my saying it outright; all I need do is hint"; while the other partner, typically the male, assumes, "We know each other so well that you will tell me what you want."

Furthermore, there are various ways of honoring interpersonal involvement, as service of two overriding human goals. These have been called, by Brown and Levinson (1978), positive and negative politeness, building on R. Lakoff's stylistic continuum from camaraderie to distance (1973) and Goffman's presentational and avoidance rituals (1967). These and other schemata recognize the universal human needs to 1) be connected to other people and 2) be left alone. Put another way, there are universal, simultaneous, and conflicting human needs for community and independence.

Linguistic choices reflect service of one or the other of these needs in various ways. The paralinguistically gross listener responses mentioned above are features in an array of devices which I have hypothesized place the signalling load (Gumperz' term) on the need for community. Other features co-occurring in the speech of many speakers of this style include fast rate of speech; fast turn-taking; preference for simultaneous speech; tendency to introduce new topics without testing the conversational waters through hesitation and other signals; persistence in introducing topics not picked up by others; storytelling; preference for stories told about personal experience and revealing emotional reaction of teller; talk about personal matters; overstatement for effect. (All of these features surfaced in the setting of a casual conversation at dinner; it would be premature to generalize for other settings). These and other features of the speech of the New Yorkers sometimes struck the Californians present as imposing, hence failing to honor their need for independence. The use of contrasting devices by the Californians led to the impression on some of the New Yorkers that they were deficient in honoring the need for community. Of course the underlying goals were not conceptualized by participants at the time. What was perceived was sensed as personality characteristics: "They're dominating," and "They're cold." Conversely, when style was shared, the conclusion was, "They're nice."

Perhaps many of these stylistic differences come down to differing attitudes toward silence. I suggest that the fast-talking style I have characterized above grows out of a desire to avoid silence, which has a negative value. Put another way, the unmarked meaning of silence, in this system, is evidence of lack of rapport. To other speakers -- for example, Athabaskan Indians, according to Basso (1972) and Scollon (1980) -- the unmarked meaning of silence is positive.

Individual and social differences

All of these parameters are intended to suggest processes that operate in signalling meaning in conversation. Analysis of cross-cultural differences is useful to make apparent processes that go unnoticed when signalling systems are shared.

An obvious question, one that has been indirectly addressed throughout the present discussion, confronts the distinction between individual and cultural differences. We need to know, for the understanding of our own lives as much as for our theoretical understanding of discourse, how much of any speaker's style -- the linguistic and paralinguistic devices signalling meaning -- are prescribed by the culture, and which are chosen freely. The answer to this seems to resemble, one level further removed, the distinction between cohesive vs. expressive features. The answer, furthermore, must lie somewhere between fixity and novelty -- a matter of choices among alternatives offered by cultural convention.

References

Basso, K. 1972. To give up on words: Silence in Western Apache culture, in P.P. Giglioli, ed., Language in social context. Penguin.

Brown, P. & S. Levinson. 1978. Universals in language usage: Politeness phenomena, in E. Goody, ed., Questions and politeness. Cambridge.

Fillmore, C. 1979. Innocence: A second idealization for linguistics. Proceedings of the fifth annual meeting of the Berkeley Linguistics Society.

Goffman, E. 1967. Interaction ritual. Doubleday.

Kay, P. 1977. Language evolution and speech style, in B. Blount & M. Sanches, eds., Sociocultural dimensions of language change. NY: Academic.

Lakoff, R. 1973. The logic of politeness, or minding your p's and q's. Papers from the ninth regional meeting of the Chicago Linguistics Society.

Scollon, R. 1980. The machine stops: Silence in the metaphor of malfunction. Paper prepared for the American Anthropological Association annual meeting.

Interactive Discourse: Influence of the Social Context
Panel Chair's Introduction

Jerry R. Hobbs
SRI International

Progress in natural language interfaces can perhaps be stimulated and directed by imagining the ideal natural language system of the future. What features (or even design philosophies) should such a system have in order to become an integral part of our work environments? What scaled-down versions of these features might be possible in the near future in "simple service systems"? These issues can be broken down into the following four questions.

1. What are the significant features of the environment in which the system will reside? The system will be one participant in an intricate information network, depending on a continually reinforced shared complex of knowledge [9]. To be an integral part of this environment, the system must possess some of the shared knowledge and perhaps must participate in its reinforcement, e.g. via explanations, [9], [2].

2. Investigations of person-person communication should tell us what person-system communication ought to be like. Face-to-face conversation is extraordinarily rich in the information that is conveyed by various means, such as gesture, body position, gaze direction [4], [8]. In addition to conveying propositional content or information, what are the principal functions that moves in conversation perform?

a. Organization of the interaction, regulation of turns [7], [1]. In the natural language dialog systems of today, each turn consists of a sentence or less. In experiments done at SRI on instruction dialogs between people over computer terminals, the instructor's turns usually involve long texts. It was discovered that the student needs a way of interrupting. That is, some sort of turn-taking mechanisms are required. What can we learn from the turn-taking mechanisms people use?

b. Orientation of the participants toward each other, including recognition [6], expressions of solidarity and indications of agreement and disagreement [3], meta-comments on the direction of the conversation [8] or the reasons for certain utterances ([9] on discourse explanations).

c. Maintenance of the channel of communication, implicit acknowledgment or verification of information conveyed [2]. Recovery from mistakes and breakdowns in communication [8], e.g. via flexibility in parsing and interpretation [2]; via explicit indications of incomprehension [2] and repairs [5]. In natural language systems of today, when the user makes a mistake and the system fails to interpret the input, the user must usually begin over again. The system cannot use whatever it did get from the mistake to aid in the interpretation of the repair. People are more efficient. What are the principal means of repair that people use, and how can they be carried over to natural language systems?

d. Building and reinforcing the mutual knowledge base, i.e. the knowledge the participants share and know they share, etc. [9]. Linking new or out-of-the-ordinary information to shared knowledge via explanations [9], [2].

e. Inferring others' goals, knowledge, abilities, focus of attention [8], [2], [4]. The system should have a model of the user and of the communication situation [8].

f. Communicating one's own goals, knowledge, abilities, focus of attention [8], [2]. Establishing and maintaining one's role, e.g. as a competent, cooperative participant (cf. [8]; [9]; [1] for the role of speech style; [4] for defense of competence). In addition to the system having a model of the user, the user will have a model of the system, determined by the nature of his interaction with it. The system should thus be tailored to convey an accurate image of what the system can do. For example, superficial politeness or fluency ("Good morning, Jerry. What can I do for you today?") is more likely to mislead the user about the system's capabilities than to ease the interaction. What the system does, via lexical choice, indirect speech acts, polite forms, etc., to maintain its role in the interaction should arise out of a coherent view of what the role is. The linguistic competence of the system is an important element of the image it conveys to the user [2].

3. When we move from face-to-face conversations to dialogs over computer terminals, the communication is purely verbal. The work done non-verbally now has to be realized verbally. How are the realizations of the above functions altered over the change of channels [6]? We know, for example, that there are more utterances showing solidarity and asking for opinions, because this is work done non-verbally face-to-face [3]. Some things that occur face-to-face (e.g. tension release, jokes) seem to be expendable over computer terminals, where each utterance costs the speaker more. The messages take longer to produce, are less transitory, and can be absorbed more carefully, so there is less asking for orientation, elaboration, and correction [3]. What devices are likely to be borrowed from related but more familiar communication frames [1]? Possible frames are letters or telephone conversations.

4. Should and how can these functions be incorporated into the ideal natural language systems of the far future and the simple service systems of the near future [2], [8]?

REFERENCES

1. Carey, J. Interactive television: A frame analysis. From M. Moss (ed.), Two-Way Cable Television: An Evaluation of Community Uses in Reading, Pennsylvania. Final report to the National Science Foundation. 1978.

2. Hayes, P. and R. Reddy. An anatomy of graceful interaction in spoken and written man-machine communication. Computer Science Department, Carnegie-Mellon University. 1979.

3. Hiltz, S. R., K. Johnson, C. Aronovitch, and M. Turoff. Face to face vs. computerized conferences: A controlled experiment. Draft final report for grant with Division of Mathematical and Computer Sciences, National Science Foundation. 1980.

4. Hobbs, J. and D. Evans. Conversation as planned behavior. Technical Note 203. SRI International. 1979.

5. Sacks, H., E. Schegloff and G. Jefferson. A simplest systematics for the organization of turn-taking for conversation. Language, Vol. 50, no. 2, 696-735. 1974.

6. Schegloff, E., G. Jefferson and H. Sacks. The preference for self-correction in the organization of repair in conversation. Language, vol. 53, no. 2, 361-382. 1977.

7. Schegloff, E. Identification and recognition in

telephone conversation openings. In G. Psathas (ed.), Everyday Language: Studies in Ethnomethodology. 23-78.

8. Thomas, J. A design-interpretation analysis of natural English with applications to man-computer interaction. International Journal of Man-Machine Studies, Vol. 10, 651-668. 1978.

9. Wynn, E. Office conversation as an information medium. Ph.D. Thesis, Department of Anthropology, University of California, Berkeley. 1979.

PARALANGUAGE IN COMPUTER MEDIATED COMMUNICATION

John Carey
Alternate Media Center
New York University

This paper reports on some of the components of person to person communication mediated by computer conferencing systems. Transcripts from two systems were analysed: the Electronic Information and Exchange System (EIES), based at the New Jersey Institute of Technology; and Planet, based at Infomedia Inc. in Palo Alto, California. The research focused upon the ways in which expressive communication is encoded by users of the medium[1].

1. INTRODUCTION

The term paralanguage is used broadly in this report. It includes those vocal features outlined by Trager (1964) as well as the prosodic system of Crystal (1969). Both are concerned with the investigation of linguistic phenomena which generally fall outside the boundaries of phonology, morphology and lexical analysis. These phenomena are the voice qualities and tones which communicate expressive feelings, indicate the age, health and sex of a speaker, modify the meanings of words, and help to regulate interaction between speakers.

Paralanguage becomes an issue in print communication when individuals attempt to transcribe (and analyse) an oral presentation, or write a script which is to be delivered orally. In addition, paralinguistic analysis can be directed towards forms of print which mimic or contain elements of oral communication. These include comic strips, novels, graffitti, and computer conferencing (see Crystal and Davy 1969).

The research reported here is not concerned with a direct comparison between face-to-face and computer mediated communication. Such a comparison is useful, e.g. it can help us to understand how one form borrows elements from the other (see section 5.), or aid in the selection of the medium which is more appropriate for a given task. However, the intent here is simpler: to isolate some of the paralinguistic features which are present in computer mediated communication and to begin to map the patterning of those features.

2. THE FRAME

Computer conferencing may be described as a frame of social activity in Goffman's terms (1974). The computer conferencing frame is characterized by an exchange of print communication between or among individuals. That is, it may involve person to person or person to group communication. The information is typed on a computer terminal, transmitted via a telephone line to a central computer where it is processed and stored until the intended receiver (also using a computer terminal and a telephone line) enters the system. The received information is either printed on paper or displayed on a television screen. The exchange can be in real time, if the users are on the system simultaneously and linked together in a common notepad. More typically, the exchange is asynchronous with several hours or a few days lapse between sending and receiving.

In all of the transcripts examined for this study, the composer of the message typed it into the system. Further, the systems were used for many purposes: simple message sending (electronic mail), task related conferencing, and fun (e.g. jokes and conferences on popular topics). Bills for usage were paid by the organizations involved, not the individuals themselves. These elements within the frame may affect the style of interaction.

One concern in frame analysis is to understand differences in a situation which make a difference. Clearly, there is a need to investigate conditions not included in this study in order to gain a broader understanding of paralinguistic usage. Among the conditions which might make a difference are: the presence of a secretary in the flow of information; usage based upon narrow task communications only; and situations where there is a direct cost to the user.

3. FEATURES

The following elements have been isolated within the transcripts and given a preliminary designation as paralinguistic features.

3.1. VOCAL SPELLING

These features include non standard spellings of words which bring attention to sound qualities. The spelling may serve to mark a regional accent or an idiosyncratic manner of speech. Often, the misspelling involves repetition of a vowel (drawl) or a final consonant (released or held consonant, with final stress). In addition, there are many examples of non standard contractions. A single contraction in a message appears to bring attention (stress) to the word. A series of contractions in a single message appears to serve as a tempo marker, indicating a quick pace in composing the message.

/biznis/

/weeeeell/

/breakkk/

/y'all/

/Miami Dade Cmty Coll Life Lab Pgm/

Figure 1. Examples of Vocal Spelling

Some of the spellings shown above can occur through a glitch in the system or an unintended error by the composer of the message. Typically, the full context helps the reader to discern if the spelling was intentional.

3.2. LEXICAL SURROGATES

Often, people use words to describe their "tone of voice" in the message. This may be inserted as a parenthetical comment within a sentence, in which case it is likely to mark that sentence alone. Alternatively, it may be located at the beginning or end of a message. In these instances, it often provides a tone for the entire message.

In addition, vocal segregates (e.g. uh huh, hmmm, yuk yuk) are written commonly within the body of texts.

1. The research was supported by DHEW Grant No. 54-P-71362/2/2-01

/What was decided? I like the idea, but then again, it was mine (she said blushingly)./

/Boo, boo Horror of horrors! t165 DOESN'T seem to cure all the problems involved in transmitting files./

Figure 2. Examples of Lexical Surrogates

3.3. SPATIAL ARRAYS

Perhaps the most striking feature of computer conferencing is the spatial arrangement of words. While some users borrow a standard letter format, others treat the page space as a canvass on which they paint with words and letters, or an advertisement layout in which they are free to leave space between words, skip lines, and paragraph each new sentence.

Some spatial arrays are actual graphics: arrangements of letters to create a picture. Hiltz and Turoff (1978) note the heavy use of graphics at Christmas time, when people send greeting cards through the conferencing system. In day to day messaging, users often leave space between words (indicating pause, or setting off a word or phrase), run words together (quickening of tempo, onomatopoeic effect), skip lines within a paragraph (to setoff a word, phrase or sentence), and create paragraphs to lend visual support to the entire message or items within it. In addition, many messages contain headlines, as in newspaper writing.

/One of our units here just makes an awfulhowling noise./

/AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
MMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMMM
OOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOO
SSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSS/

/$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$$
When the next bill comes in from
EIES/Telenet, you may also be interested/

Figure 3. Examples of Spatial Arrays

3.4. MANIPULATION OF GRAMMATICAL MARKERS

Grammatical markers such as capitalization, periods, commas, quotation marks, and parentheses are manipulated by users to add stress, indicate pause, modify the tone of a lexical item and signal a change of voice by the composer. For example, a user will employ three exclamation marks at the end of a sentence to lend intensity to his point. A word in the middle of a sentence (or one sentence in a message) will be capitalized and thereby receive stress. A series of dashes between syllables of a word can serve to hold the preceding syllable and indicate stress upon it or the succeeding syllable. Parentheses and quotation marks are used commonly to indicate that the words contained within them are to be heard with a different tone than the rest of the message. A series of periods are used to indicate pause, as well as to indicate internal and terminal junctures. For example, in some messages, composers do not use commas. At points where a comma is appropriate, three periods are employed. At the end of the sentence, several periods (the number can vary from 4 to more than 20) are used. This system indicates to the reader both the grammatical boundary and the length of pause between words.

The Electronic Information and Exchange System employs some of these grammatical marker manipulations in the interface between user and system. For example, they instruct a user to respond with question marks when he does not know what to do at a command point. One question mark indicates "I don't understand what EIES wants here," and will yield a brief explanation from the system. Two question marks indicate "I am <u>very</u> confused" and yield a longer explanation. Three question marks indicate "I am totally lost" and put the user in direct touch with the system monitor.

/Welcome Aboard!!!!/

/This background is VERY important, since it makes many people (appropriately, I think) aware about idea./

/THERE IS STILL SOME CONFUSION ON DATES FOR PHILADELPHIA. MIKE AND I ARE PERPLEXED!?/

/At this point, I think we should include a BROAD range of ideas -- even if they look unworkable./

/Paul...three quick points......first...the paper/

Figure 4. Manipulation of Grammatical Markers

3.5. MINUS FEATURES

The absence of certain features or expected work in composition may also lend a tone to the message. For example, a user may not correct spelling errors or glitches introduced by the system. Similarly, he may pay no attention to paragraphing or capitalization. The absence of such features, particularly if they are clustered together in a single message, can convey a relaxed tone of familiarity with the receiver or quickness of pacing (e.g. when the sender has a lot of work to do and must compose the message quickly).

4. <u>PATTERNING OF FEATURES</u>

It can be noted, first, that some features mark a short syllabic or polysyllabic segment (e.g. capitalization, contraction, and vocal segregates), while others mark full sentences or the entire message (e.g. a series of exclamation points, letter graphics, or an initial parenthetical comment). Second, it is revealing that many of these features have an analogic structure: in some manner, they are like the tone they represent. For example, a user may employ more or fewer periods, more or fewer question marks to indicate degrees of pause or degrees of perplexity. Paralanguage in everyday conversation is highly analogic and represents feelings, moods and states of health which do not (apparently) lend themselves to the digital structure of words.

Paralinguistic features in computer conferencing occur, often, at points of <u>change</u> in a message: change of pace, change of topic, change of tone. In addition, many of the features rely upon a contrastive structure to communicate meaning. That is, a message which is typed in all caps does not communicate greater intensity or stress. Capitalization must occur contrastively over one or two words in an otherwise normal sentence or over one or two sentences in a message which contains some normal capitalization.

Most paralinguistic features can have more than one meaning. Reviewed in isolation, a feature might indicate a relaxed tone, an intimate relation with the receiver, or simply sloppiness in composition. Readers must rely upon the surrounding context (both words and other paralinguistic features) to narrow the range of possible meanings.

The intended receiver of a message, as well as an outsider who attempts to analyse transcripts, must cope with the interpretation of paralinguistic features. Initially, the reader must distinguish glitches in the system and unintended typing errors from intentional use of repetition, spacing, etc. Subsequently, the reader must examine the immediate context of the feature and compare the usage with similar patterns in the same message, in other messages by the composer, and/or in other messages by the general population of users.

5. DEVELOPMENT OF A CODE

The findings presented in this study are taken from a limited set of contexts. For this reason, they must be regarded as a first approximation of paralinguistic code structure in computer conferencing. Moreover, the findings do not suggest that a clear code exists for the community of users. Rather, the code appears to be in a stage of development and learning.

The study has helped to define some differences among users which appear to make a difference in the paralinguistic features they employ. In the corpus of transcripts examined, usage varied between new and experienced participants, as well as between infrequent and frequent participants. Generally, experienced and frequent participants employed more paralinguistic features. However, idiosyncratic patterns appear to be more important in determining usage. The findings serve more to define questions for subsequent study than to provide answers about user variations.

In addition, it is clear that the characteristics of the computer terminals (TI 745s, primarily), as well as system characteristics, provided many of the components or "bricks" with which paralinguistic features were constructed. For example, the repeat key on the terminal allowed users to create certain forms of graphics. Also, star keys, dollar signs, colons and other available keys were employed to communicate paralinguistic information. System terms to describe a mode of operation (e.g. notepad, scratchpad, message, conference) may also influence development of a code of usage by suggesting a more formal or informal exchange.

Finally, it may be noted that early in their usage, some participants appeared to borrow formats from other media with which they were familiar (e.g. business letters, telegrams, and telephone conversations). Over time, patterns of usage converged somewhat. However, idiosyncratic variation remained strong.

6. CONCLUSION

A few conclusions can be drawn from this study. First, the presence of paralinguistic features in computer conferencing and the effort by users to communicate more information than can be carried by the words themselves, suggest that people feel it is important to be able to communicate tonal and expressive information. Second, it is not easy to communicate this information. Users must work in computer conferencing to communicate information about their feelings and state of health which naturally accompanies speech. While there does not appear to be a unified and identifiable code of paralinguistic features within conferencing systems or among users of the systems, the collective behavior of participants may be creating one.

REFERENCES

Crystal, David Prosodic Systems and Intonation in English. Cambridge: Cambridge University Press 1969.

Crystal, David and Davy, Derek Investigating English Style. Bloomington: Indiana University Press 1969.

Goffman, Erving Frame Analysis. New York: Harper and Row 1974.

Hiltz, Starr Roxanne and Turoff, Murray The Network Nation. Reading, Massachusetts: Addison-Wesley 1978.

Trager, George "Paralanguage: A First Approximation," in Dell Hymes (ed.) Language in Culture and Society. New York: Harper and Row 1964.

Expanding the Horizons of Natural Language Interfaces

Phil Hayes
Computer Science Department, Carnegie-Mellon University
Pittsburgh, PA 15213, USA

Abstract

Current natural language interfaces have concentrated largely on determining the literal "meaning" of input from their users. While such decoding is an essential underpinning, much recent work suggests that natural language interfaces will never appear cooperative or graceful unless they also incorporate numerous non-literal aspects of communication, such as robust communication procedures.

This paper defends that view, but claims that direct imitation of human performance is not the best way to implement many of these non-literal aspects of communication; that the new technology of powerful personal computers with integral graphics displays offers techniques superior to those of humans for these aspects, while still satisfying human communication needs. The paper proposes interfaces based on a judicious mixture of these techniques and the still valuable methods of more traditional natural language interfaces.

1. Introduction

Most work so far on natural language communication between man and machine has dealt with its literal aspects. That is, natural language interfaces have implicitly adopted the position that their user's input encodes a request for information or action, and that their job is to decode the request, retrieve the information, or perform the action, and provide appropriate output back to the user. This is essentially what Thomas [24] calls the *Encoding-Decoding model of conversation*.

While literal interpretation is a basic underpinning of communication, much recent work in artificial intelligence, linguistics, and related fields has shown that it is far from the whole story in human communication. For example, appropriate interpretation of an utterance depends on assumptions about the speaker's intentions, and conversely, the speaker's goals influence what is said (Hobbs [13], Thomas [24]). People often make mistakes in speaking and listening, and so have evolved conventions for effecting repairs (Schegloff et al. [20]). There must also be a way of regulating the turns of participants in a conversation (Sacks et al. [19]). This is just a sampling of what we will collectively call *non-literal aspects of communication*.

The primary reason for using natural language in man-machine communication is to allow the user to express himself naturally, and without having to learn a special language. However, it is becoming clear that providing for natural expression means dealing with the non-literal as well as the literal aspects of communication, that the ability to interpret natural language literally does not in itself give a man-machine interface the ability to communicate naturally. Some work on incorporating these non-literal aspects of communication into man-machine interfaces has already begun ([6, 8, 9, 15, 21, 25]).

The position I wish to stress in this paper is that natural language interfaces will never perform acceptably unless they deal with the non-literal as well as the literal aspects of communication, that without the non-literal aspects, they will always appear uncooperative, inflexible, unfriendly, and generally stupid to their users, leading to irritation, frustration, and an unwillingness to continue to be a user.

This position is coming to be held fairly widely. However, I wish to go further and suggest that, in building non-literal aspects of communication into natural-language interfaces, we should aim for the most effective type of communication rather than insisting that the interface model human performance as exactly as possible. I believe that these two aims are not necessarily the same, especially given certain new technological trends discussed below.

Most attempts to incorporate non-literal aspects of communication into natural language interfaces have attempted to model human performance as closely as possible. The typical mode of communication in such an interface, in which system and user type alternately on a single scroll of paper (or scrolled display screen), has been used as an analogy to normal spoken human conversation in which communication takes place over a similar half-duplex channel, i.e. a channel that only one party at a time can use without danger of confusion.

Technology is outdating this model. The nascent generation of powerful personal computers (e.g. the ALTO [23] or PERQ [18]) equipped with high-resolution bit-map graphics display screens and pointing devices allow the rapid display of large quantities of information and the maintenance of several independent communication channels for both output (division of the screen into independent windows, highlighting, and other graphics techniques), and input (direction of keyboard input to different windows, pointing input). I believe that this new technology can provide highly effective, natural language-based, communication between man and machine, but only if the half-duplex style of interaction described above is dropped. Rather than trying to imitate human conversation directly, it will be more fruitful to use the capabilities of this new technology, which in some respects exceed those possessed by humans, to achieve the same ends as the non-literal aspects of normal human conversation. Work by, for instance, Carey [3] and Hiltz [12] shows how adaptable people are to new communication situations, and there is every reason to believe that people will adapt well to an interaction in which their communication *needs* are satisfied, even if they are satisfied in a different way than in ordinary human conversation.

In the remainder of the paper I will sketch some human communication needs, and go on to suggest how they can be satisfied using the technology outlined above.

2. Non-Literal Aspects of Communication

In this section we will discuss four human communication needs and the non-literal aspects of communication they have given rise to:

- non-grammatical utterance recognition
- contextually determined interpretation
- robust communication procedures
- channel sharing

The account here is based in part on work reported more fully in [8, 9].

Humans must deal with **non-grammatical utterances** in conversation simply because people produce them all the time. They arise from various sources: people may leave out or swallow words; they may start to say one thing, stop in the middle, and substitute something else; they may interrupt themselves to correct something they have just said, or they may simply make errors of tense, agreement, or vocabulary. For a combination of these and other reasons, it is very rare to see three consecutive grammatical sentences in ordinary conversation.

Despite the ubiquity of ungrammaticality, it has received very little attention in the literature or from the implementers of natural-language interfaces. Exceptions include PARRY [17], COOP [14], and interfaces produced by the LIFER [11] system. Additional work on parsing ungrammatical input has been done by Weischedel and Black [25], and

Kwasny and Sandheimer [15]. As part of a larger project on user interfaces [1], we (Hayes and Mouradian [7]) have also developed a parser capable of dealing flexibly with many forms of ungrammaticality

Perhaps part of the reason that flexibility in parsing has received so little attention in work on natural language interfaces is that the input is typed, and so the parsers used have been derived from those used to parse written prose. Speech parsers (see for example [10] or [26]) have always been much more flexible. Prose is normally quite grammatical simply because the writer has had time to make it grammatical. The typed input to a computer system is produced in "real time" and is therefore much more likely to contain errors or other ungrammaticalities

The listener at any given turn in a conversation does not merely decode or extract the inherent "meaning" from what the speaker said. Instead he interprets the speaker's utterance in the light of the total available context (see for example, Hobbs [13], Thomas [24] or Wynn [27]). In cooperative dialogues, and computer interfaces normally operate in a cooperative situation, this **contextually determined interpretation** allows the participants considerable economies in what they say substituting pronouns or other anaphoric forms for more complete descriptions, not explicitly requesting actions or information that they really desire, omitting participants from descriptions of events, and leaving unsaid other information that will be "obvious" to the listener because of the context shared by speaker and listener. In less cooperative situations, the listener's interpretations may be other than the speaker intends, and speakers may compensate for such distortions in the way they construct their utterances.

While these problems have been studied extensively in more abstract natural language research (for just a few examples see [4, 5, 16]), little attention has been paid to them in more applied language work. The work of Grosz [6] and Sidner [21] on focus of attention and its relation to anaphora and ellipsis stand out here, along with work done in the COOP [14] system on checking the presuppositions of questions with a negative answer. In general, contextual interpretation covers most of the work in natural language processing, and subsumes numerous currently intractable problems. It is only tractable in natural language interfaces because of the tight constraints provided by the highly restricted worlds in which they operate.

Just as in any other communication across a noisy channel, there is always a basic question in human conversation of whether the listener has received the speaker's utterance correctly. Humans have evolved **robust communication** conventions for performing such checks with considerable, though not complete, reliability, and for correcting errors when they occur (see Schegloff [20]). Such conventions include: the speaker assuming an utterance has been heard correctly unless the reply contradicts this assumption or there is no reply at all; the speaker trying to correct his own errors himself; the listener incorporating his assumptions about a doubtful utterance into his reply; the listener asking explicitly for clarification when he is sufficiently unsure.

This area of robust communication is perhaps the non-literal aspect of communication most neglected in natural language work. Just a few systems such as LIFER [11] and COOP [14] have paid even minimal attention to it. Interestingly, it is perhaps the area in which the new technology mentioned above has the most to offer as we shall see.

Finally, the spoken part of a human conversation takes place over what is essentially a single shared channel. In other words, if more than one person talks at once, no one can understand anything anyone else is saying. There are marginal exceptions to this, but by and large reasonable conversation can only be conducted if just one person speaks at a time. Thus people have evolved conventions for **channel sharing** [19], so that people can take turns to speak. Interestingly, if people are put in new communication situations in which the standard turn-taking conventions do not work well, they appear quite able to evolve new conventions [3]

As noted earlier, computer interfaces have sidestepped this problem by making the interaction take place over a half-duplex channel somewhat analogous to the half duplex channel inherent in speech, i.e. alternate turns at typing on a scroll of paper (or scrolled display screen). However, rather than providing flexible conventions for changing turns, such interfaces typically brook no interruptions while they are typing, and then when they are finished insist that the user type a complete input with no feedback (apart from character echoing), at which point the system then takes over the channel again

In the next section we will examine how the new generation of interface technology can help with some of the problems we have raised

3. Incorporating Non-Literal Aspects of Communication into User Interfaces

If computer interfaces are ever to become cooperative and natural to use, they must incorporate non-literal aspects of communication. My main point in this section is that there is no reason they should incorporate them in a way directly imitative of humans, so long as they are incorporated in a way that humans are comfortable with; direct imitation is not necessary. Indeed, direct imitation is unlikely to produce satisfactory interaction. Given the present state of natural language processing and artificial intelligence in general, there is no prospect in the forseeable future that interfaces will be able to emulate human performance, since this depends so much on bringing to bear larger quantities of knowledge than current AI techniques are able to handle. Partial success in such emulation is only likely to raise false expectations in the mind of the user, and when these expectations are inevitably crushed, frustration will result. However, I believe that by making use of some of the new technology mentioned earlier, interfaces can provide very adequate substitutes for human techniques for non-literal aspects of communication, substitutes that capitalize on capabilities of computers that are not possessed by humans, but that nevertheless will result in interaction that feels very natural to a human

Before giving some examples, let us review the kind of hardware I am assuming. The key item is a bit-map graphics display capable of being filled with information very quickly. The screen can be divided into independent windows to which the system can direct different streams of output independently. Windows can be moved around on the screen, overlapped, and popped out from under a pile of other windows. The user has a pointing device with which he can position a cursor to arbitrary points on the screen, plus, of course, a traditional keyboard. Such hardware exists now and will become increasingly available as powerful personal computers such as the PERQ [18] or LISP machine [2] come onto the market and start to decrease in price. The examples of the use of such hardware which follow are drawn in part from our current experiments in user interface research [1, 7] on similar hardware.

Perhaps the aspect of communication that can receive the most benefit from this type of hardware is robust communication. Suppose the user types a non-grammatical input to the system which the system's flexible parser is able to recognize if, say, it inserts a word and makes a spelling correction. Going by human convention the system would either have to ask the user to confirm explicitly if its correction was correct, to cleverly incorporate its assumption into its next output, or just to assume the correction without comment. Our hypothetical system has another option: it can alter what the user just typed (possibly highlighting the words that it changed). This achieves the same effect as the second option above, but substitutes a technological trick for human intelligence

Again, if the user names a person, say "Smith", in a context where the system knows about several Smiths with different first names, the human options are either to incorporate a list of the names into a sentence (which becomes unwieldy when there are many more than three alternatives) or to ask for the first name without giving alternatives. A third alternative, possible only in this new technology, is to set up a window on the screen

with an initial piece of text followed by a list of alternatives (twenty can be handled quite naturally this way) The user is then free to point at the alternative he intends, a much simpler and more natural alternative than typing the name, although there is no reason why this input mode should not be available as well in case the user prefers it

As mentioned in the previous section, contextually based interpretation is important in human conversation because of the economies of expression it allows There is no need for such economy in an interface's output, but the human tendency to economy in this matter is something that technology cannot change The general problem of keeping track of focus of attention in a conversation is a difficult one (see, for example, Grosz [6] and Sidner [22]), but the type of interface we are discussing can at least provide a helpful framework in which the current focus of attention can be made explicit Different foci of attention can be associated with different windows on the screen, and the system can indicate what it thinks is the current focus of attention by, say, making the border of the corresponding window different from all the rest Suppose in the previous example that at the time the system displays the alternative Smiths, the user decides that he needs some other information before he can make a selection He might ask for this information in a typed request, at which point the system would set up a new window, make it the focused window, and display the requested information in it At this point, the user could input requests to refine the new information, and any anaphora or ellipsis he used would be handled in the appropriate context.

Representing contexts explicitly with an indication of what the system thinks is the current one can also prevent confusion The system should try to follow a user's shifts of focus automatically as in the above example However, we cannot expect a system of limited understanding always to track focus shifts correctly, and so it is necessary for the system to give explicit feedback on what it thinks the shift was Naturally, this implies that the user should be able to change focus explicitly as well as implicitly (probably by pointing to the appropriate window).

Explicit representation of foci can also be used to bolster a human's limited ability to keep track of several independent contexts In the example above, it would not have been hard for the user to remember why he asked for the additional information and to return and make the selection after he had received that information With many more than two contexts however, people quickly lose track of where they are and what they are doing Explicit representation of all the possibly active tasks or contexts can help a user keep things straight

All the examples of how sophisticated interface hardware can help provide non-literal aspects of communication have depended on the ability of the underlying system to produce possibly large volumes of output rapidly at arbitrary points on the screen In effect, this allows the system multiple output channels independent of the user's typed input, which can still be echoed even while the system is producing other output. Potentially this frees interaction over such an interface from any turn taking discipline In practice, some will probably be needed to avoid confusing the user with too many things going on at once, but it can probably be looser than that found in human conversations.

As a final point, I should stress that natural language capability is still extremely valuable for such an interface While pointing input is extremely fast and natural when the object or operation that the user wishes to identify is on the screen, it obviously cannot be used when the information is not there Hierarchical menu systems, in which the selection of one item in a menu results in the display of another more detailed menu, can deal with this problem to some extent but the descriptive power and conceptual operators of natural language (or an artificial language with similar characteristics) provide greater flexibility and range of expression If the range of options is large, but well discriminated, it is often easier to specify a selection by description than by pointing, no matter how cleverly the options are organized

4. Conclusion

In this paper, I have taken the position that natural language interfaces to computer systems will never be truly natural until they include non-literal as well as literal aspects of communication Further, I claimed that in the light of the new technology of powerful personal computers with integral graphics displays, the best way to incorporate these non-literal aspects was not to imitate human conversational patterns as closely as possible but to use the technology in innovative ways to perform the same function as the non-literal aspects of communication found in human conversation

In any case, I believe the old-style natural language interfaces in which the user and system take turns to type on a single scroll of paper (or scrolled display screen) are doomed The new technology can be used, in ways similar to those outlined above, to provide very convenient and attractive interfaces that do not deal with natural language. The advantages of this type of interface will so dominate those associated with the old-style natural language interfaces that continued work in that area will become of academic interest only.

That is the challenge posed by the new technology for natural language interfaces but it also holds a promise The promise is that a combination of natural language techniques with the new technology will result in interfaces that will be truly natural, flexible, and graceful in their interaction The multiple channels of information flow provided by the new technology can be used to circumvent many of the areas where it is very hard to give computers the intelligence and knowledge to perform as well as humans In short the way forward for natural language interfaces is not to strive for closer but still highly imperfect, imitation of human behaviour, but to combine the strengths of the new technology with the great human ability to adapt to communication environments which are novel but adequate for their needs.

References

1. Ball, J E. and Hayes, P. J. Representation of Task-Independent Knowledge in a Gracefully Interacting User Interface. Tech. Rept., Carnegie-Mellon University Computer Science Department, 1980.

2. Bawden, A, et al. Lisp Machine Project Report AIM 444, MIT AI Lab, Cambridge, Mass., August, 1977.

3. Carey, J "A Primer on Interactive Television" *J University Film Assoc XXX*, 2 (1978), 35-39.

4. Charniak, E C Toward a Model of Children's Story Comprehension. TR-266, MIT AI Lab, Cambridge, Mass., 1972.

5. Cullingford, R *Script Application Computer Understanding of Newspaper Stories* Ph.D Th., Computer Science Dept., Yale University, 1978

6. Grosz, B J. The Representation and Use of Focus in a System for Understanding Dialogues Proc. Fifth Int. Jt. Conf. on Artificial Intelligence, MIT, 1977, pp 67-76.

7. Hayes, P J and Mouradian, G V Flexible Parsing Proc. of 18th Annual Meeting of the Assoc for Comput. Ling., Philadelphia, June, 1980.

8. Hayes, P J and Reddy, R Graceful Interaction in Man-Machine Communication Proc Sixth Int Jt Conf. on Artificial Intelligence, Tokyo, 1979, pp 372-374

9. Hayes, P J and Reddy, R An Anatomy of Graceful Interaction in Man-Machine Communication Tech report, Computer Science Department, Carnegie-Mellon University, 1979.

10. Hayes-Roth, F., Erman, L. D., Fox, M., and Mostow, D. J. Syntactic Processing in HEARSAY-II. Speech Understanding Systems: Summary of Results of the Five-Year Research Effort at Carnegie-Mellon University, Carnegie-Mellon University Computer Science Department, 1976.

11. Hendrix, G. G. Human Engineering for Applied Natural Language Processing. Proc. Fifth Int. Jt. Conf. on Artificial Intelligence, MIT, 1977, pp. 183-191.

12. Hiltz, S. R., Johnson, K., Aronovitch, C., and Turoff, M. Face to Face vs. Computerized Conferences: A Controlled Experiment. unpublished mss.

13. Hobbs, J. R. Conversation as Planned Behavior. Technical Note 203, Artificial Intelligence Center, SRI International, Menlo Park, Ca., 1979.

14. Kaplan, S. J. *Cooperative Responses from a Portable Natural Language Data Base Query System*. Ph.D. Th., Dept. of Computer and Information Science, University of Pennsylvania, Philadelphia, 1979.

15. Kwasny, S. C. and Sondheimer, N. K. Ungrammaticality and Extra-Grammaticality in Natural Language Understanding Systems. Proc. of 17th Annual Meeting of the Assoc. for Comput. Ling., La Jolla, Ca., August, 1979, pp. 19-23.

16. Levin, J. A., and Moore, J. A. "Dialogue Games: Meta-Communication Structures for Natural Language Understanding." *Cognitive Science 1*, 4 (1977), 395-420.

17. Parkison, R. C., Colby, K. M., and Faught, W. S. "Conversational Language Comprehension Using Integrated Pattern-Matching and Parsing." *Artificial Intelligence 9* (1977), 111-134.

18. PERQ. Three Rivers Computer Corp., 160 N. Craig St., Pittsburgh, PA 15213. .

19. Sacks, H., Schegloff, E. A., and Jefferson, G. "A Simplest Semantics for the Organization of Turn-Taking for Conversation." *Language 50*, 4 (1974), 696-735.

20. Schegloff, E. A., Jefferson, G., and Sacks, H. "The Preference for Self-Correction in the Organization of Repair in Conversation." *Language 53*, 2 (1977), 361-382.

21. Sidner, C. L. A Progress Report on the Discourse and Reference Components of PAL. A. I. Memo. 468, MIT A. I. Lab., 1978.

22. Sidner, C. L. Towards a Computational Theory of Definite Anaphora Comprehension in English Discourse. TR 537, MIT AI Lab, Cambridge, Mass., 1979.

23. Thacker, C.P., McCreight, E.M., Lampson, B.W., Sproull, R.F., and Boggs, D.R. Alto: A personal computer. In *Computer Structures: Readings and Examples*, McGraw-Hill, 1980. Edited by D. Siewiorek, C.G. Bell, and A. Newell, second edition, in press.

24. Thomas, J. C. "A Design-Interpretation of Natural English with Applications to Man-Computer Interaction." *Int. J. Man Machine Studies 10* (1978), 651-668.

25. Weischedel, R. M. and Black, J. Responding to Potentially Unparseable Sentences. Tech. Rept. 79/3, Dept. of Computer and Information Sciences, University of Delaware, 1979.

26. Woods, W. A., Bates, M., Brown, G., Bruce, B., Cook, C., Klovstad, J., Makhoul, J., Nash-Webber, B., Schwartz, R., Wolf, J., and Zue, V. Speech Understanding Systems - Final Technical Report. Tech. Rept. 3438, Bolt, Beranek, and Newman, Inc., 1976.

THE PROCESS OF COMMUNICATION IN FACE TO FACE VS. COMPUTERIZED CONFERENCES; A CONTROLLED EXPERIMENT USING BALES INTERACTION PROCESS ANALYSIS

Starr Roxanne Hiltz, Kenneth Johnson, and Ann Marie Rabke
Upsala College

INTRODUCTION

A computerized conference (CC) is a form of communication in which participants type into and read from a computer terminal. The participants may be on line at the same time--termed a "synchrononous" conference, or may interact asynchronously. The conversation is stored and mediated by the computer.

How does this form of communication change the process and outcome of group discussions, as compared to the "normal" face to face (FtF) medium of group discussion, where participants communicate by talking, listening and observing non-verbal behavior, and where there is no lag between the sending and receipt of communication signals? This paper briefly summarizes the results of a controlled laboratory experiment designed to quantify the manner in which conversation and group decision making varies between FtF and CC. Those who wish more detail are referred to the literature review which served as the basis for the design of the experiment (Hiltz, 1975) and to the full technical report on the results (Hiltz, Johnson, Aronovitch, and Turoff, 1980). This paper is excerpted from a longer paper on the analysis of communications process in the two media and their correlates (Hiltz, Johnson and Rabke, 1980).

OVERVIEW OF THE EXPERIMENT

The chief independent variable of interest is the impact of computerized conferencing as a communications mode upon the process and outcome of group decision making, as compared to face-to-face discussions. Two different types of tasks were chosen, and group size was set at five persons. The subjects were Upsala College undergraduate, graduate and continuing education students. The communications process or profile was quantified using Bales Interaction Process Analysis (see Bales, 1950).

In computerized conferencing, each participant is physically alone with a computer terminal attached to a telephone. In order to communicate, he or she types entries into the terminal and reads entries sent by the other participants, rather than speaking and listening. Entering input and reading output may be done totally at the pace and time chosen by each individual. Conceivably, for instance, all group members could be entering comments simultaneously. Receipt of messages from others is at the terminal print speed of 30 characters per second.

Even when all five participants are on-line at the same time, there is considerable lag in a computer conference between the time a discussant types in a comment, and when a response to that comment is received. First, each of the other participants must finish what they are typing at the time; then they read the waiting item; then they may type in a response; then the author of the original comment must finish his or her typing of a subsequent item and print and read the response. There is thus a definite "asynchronous" quality even to "synchronous" computer conferences. As a result, computer conferences often develop several simultaneous threads of discussion that are being discussed concurrently, whereas face to face discussions tend to focus on one single topic at a time and then move on to subsequent topics. (See Hiltz and Turoff, 1978, for a complete description of CC as a mode of communication).

A variable of secondary interest is problem type. Much experimental literature indicates that the nature of the problem has a great deal to do with group performance. One type of problem that we used is the human relations case as developed by Bales. These are medium complex, unsettled problems that have no specific "correct" answer. The second type was a "scientific" ranking problem (requiring no specific expertise), which has a single correct solution plus measurable degrees of how nearly correct a group's answer may be. The ranking problem, "Lost in the Arctic", was adapted for administration over a conferencing system by permission of its originators (See Eady and Lafferty).

The experiments thus had a 2 x 2 factorial design (see figure one). The factors were mode of communication (face-to-face vs. computerized conference) and problem type (human relations vs. a more "scientific" ranking problem with a correct answer). These factors constituted the "independent variables." Each problem-mode condition included a total of eight groups.

Figure 1

Design of the Experiment
Two by Two Factorial with Repeated Measures:
Blocks of Four

	Task Type A	Task Type B
	Groups	
Face-to-Face	4	4
Computerized Conference	4	4

BACKGROUND: THE BALES EXPERIMENTS AND INTERACTION PROCESS ANALYSIS

Working at the Laboratory of Social Relations at Harvard, Bales and his colleagues developed a set of categories and procedures for coding the interaction in small face-to-face decision-making groups which became very widely utilized and generated a great deal of data about the nature of communication and social processes within such groups.

Coding of the communications interaction by Interaction Process Analysis involves noting who makes a statement or non-verbal participation (such as nodding agreement); to whom the action was addressed; and into which of twelve categories the action best fits. These categories are listed in subsequent tables and explained below. The distribution of communications units among the twelve categories constituted one of the main dependent variables for this experiment. We expected significant differences associated with mode of communication. We also expected some differences associated with task type. We did not feel that we had enough information to predict the directions of these differences. For almost every category, we could think of some arguments that would lead to a prediction that the category would be "higher" in CC, and some reasons why it might be lower.

METHOD

The number of Bales units per face to face group was much greater than the number for a cc group. Therefore, each individual and group was transformed to a percentage distribution among the twelve categories. Then statistical tests were performed to determine if there were any significant differences in IPA distributions associated with mode of communication, problem, order of problem, and the interaction among these variables in relation to the percentage distribution for each of the Bales categories.

There are many different ways in which the percentages could be computed. To take full advantage of the design, we computed the percentage distribution for each individual, in each condition. Thus, we actually have the Bales distributions for each of 80 individuals in a face to face conference, and in a computerized conference.

The mode of analysis was a two by two factorial nested design. If there was no significant group effect, then the error terms could be "pooled", meaning we could use the 80 observations as independent observations for statistical test purposes. We also performed a non-parametric test on the data for each Bales category, which gave us similar results.

DIFFERENCES ASSOCIATED WITH COMMUNICATION MODE

Two of the detailed analysis of variance tables on which the summary here is based are included as an Appendix. Note that the analyses were first performed separately for the two problems, using communication mode as the independent variable. For each problem, we tested the significance of mode of communication, order (whether it was the first or second problem solved by the group), and the interaction between mode and order.

Listed in figures two and three is a summary of the statistical results of the 24 analyses of variance which examined observed differences between communication modes for each of the two tasks. The first two columns show the mean percentage of communications in each category. For example, in the first table, results for Forest Ranger, the first column shows that on the average less than 1% of an individual's communications were verbally "showing solidarity", but in CC, 3.22% fell into this category. The third column shows that the results for the 16 groups in the nested factorial design were significant at the .005 level, meaning that the probability of the observed differences occuring by chance in a sample this size is one in 200. The fourth column shows the level of significance if the group was not a significant variable and the observations could be pooled, with the 80 individuals treated as independent observations. In this case, group was significant, so the pooled analysis could not be done.

In looking at these data, there is an apparent coding problem. Even for the Forest Ranger problem, face to face, we obtained a somewhat different distribution of coding than did persons coding problem discussions such as this who were directly trained by Bales. (See Bales and Borgatta, 1955, p. 400 for the complete distributions). Our coding has 20% more of the statements classified as "giving opinions" than Bales and Borgatta code, and correspondingly lower percentages in all of the other categories. This means that our results cannot be directly compared to those of other investigators, since apparently the training for coding interpreted many more statements as representing some sort of analysis or opinion than "should" be there, according to the distributions obtained for similar studies by Bales and his colleagues. (Other possible explanations are that Upsala College has produced an unusually opinionated and analytic set of students or that the effect of pre-experimental training in cc raises opinion giving even in subsequent FtF discussions.) It does not affect the comparisons among problems and modes for this study, since all of the coders were coding the data with the same guidelines and interpretations. In the majority of cases, the same pair of coders coded both the CC and FtF condition for the same group. In any case, the seven individuals who did the coding had been trained to an acceptable level of reliability.

Figure 2

Summary of IPA Results for
Forest Ranger by
Mode of Communication and Order

Bales Category	Average FTF	Average CC	P Significance By Group	P Significance Pooled
Shows:				
Solidarity	.79	3.22	.005	GS
Tension Release	3.98	.83	.0005	.0005
Agreement	13.19	4.79	.0005	.0005
Gives:				
Suggestions	4.70	9.21	.10	.10
Opinion	54.21	53.92	X	X
Orientation	12.81	16.10	.10	.02
Asks for:				
Orientation	3.27	1.58	.05	GS
Opinion	2.88	5.36	.01	.01
Suggestions	.30	.62	.25	.20
Shows:				
Disagreement	4.85	2.39	.05	.05
Tension:	.81	2.16	.05	.01
Problem 1st	.28	1.68		
Problem 2nd	1.33	2.64		
Antagonism:	.75	1.67	X	X

GS = Group significant cannot pool by individual

Figure 3

Summary of IPA Results for
Arctic by
Mode of Communication and Order

Bales Category	Average FTF	Average CC	P Significance By Group	P Significance Pooled
Shows:				
Solidarity	1.66	2.44	.10	.05
Tension Release	7.70	1.60	.0005	.0005
Agreement	13.35	6.82	.01	GS
Gives:				
Suggestions	3.56	4.89	.20	.10
Problem 1st	2.95	6.17		
Problem 2nd	4.17	3.61		
Opinion	42.99	57.80	.005	GS
Orientation	14.58	11.81	.25	GS
Asks for:				
Orientation	3.72	1.62	.025	.0005
Opinion	5.15	7.46	.20	GS
Suggestions	1.14	.58	X	GS

Shows:				
Disagreement	3.51	2.46	X	GS
Tension:	1.52	.64	.025	.005
Antagonism:	1.11	1.86	X	GS
Problem 1st	.77	.73		
Problem 2nd	1.45	3.00		

GS = Group significant cannot pool by individual

DISCUSSION OF THE RESULTS

The twelve categories in Bales Interaction Process Analysis can be combined into four main functional areas. Categories 1-3 and 10-12 are the "social-emotional" functions, oriented towards internal group process. The first three are called "social-emotional positive", while 10-12 are "negative". Categories 7-9 are "Task oriented", giving answers or contributions to solving the problem faced by the group, and categories 4-6 are varieties of "asking questions" in the task oriented area.

It will be noted, by way of further introduction, that there are some very strong differences in the profiles, even in the same medium, depending upon the type of task faced by the group, and that there is some interaction between task type and medium. For example, more tension was shown in the arctic problem in the CC condition; more in the Forest Ranger problem in the FTF condition.

We will take each of the categories, describing more fully what is included in them, and then discuss the extent to which there appear to be significant differences between the media in the relative prevalence of communications of that type. We will also try to explain the possible reasons for or implications of significant differences that are discovered.

1. "Shows solidarity, raises other's status, gives help, reward"

Included in this category are initial and responsive acts of active solidarity and affection, such as saying "hello" and making friendly or congenial remarks to "break the ice"; praising or encouraging the other(s); giving support or sympathy or offers of assistance; urging harmony and cooperation. These are all overt attempts to improve the solidarity of the group.

Note that there is a significantly greater amount of "showing solidarity" in computerized conferencing. This is probably because much of the behavior of this type in a face to face situation is non-verbal, such as smiling in a friendly manner while nodding encouragement. Non verbal acts in this category are not codable from the tapes of the discussions. In the CC condition, however, the participants realize that they must put such things into words.

Another possible explanation is that the greater tendency towards overt, explicit showing of solidarity is an attempt to compensate for the perceived coldness and impersonality of the medium.

2. "Shows Tension Release, jokes, laughs, shows satisfaction"

This includes expressions of pleasure or happiness, making friendly jokes or kidding remarks, laughing.

There was significantly more tension release overtly expressed in the face to face groups. Much of this was waves of laughter, particularly in the arctic problem. The participants did not put this into words in the conference when typing. Observing them, however, there was much private laughter and verbal expressions showing "tension release", but these do not appear in the transcript. It is part of the private "letting down of face" that occurs but is not communicated through the computer.

3. "Agrees, shows passive acceptance, understands, concurs, complies"

This occurs as concurrence in a proposed course of action or carrying out of any activity which has been requested by others. There is significantly more agreement overtly expressed in face to face conferences than in computerized conferences. We suspect that this is related to the pressure to conform created by non-verbal behavior and the physical presence of the other group members. In any case, it is undoubtedly related to the greater difficulty of CC groups in reaching total consensus.

4. "Gives Suggestion, direction, implying autonomy for other"

Includes giving suggestions about the task or suggesting concrete actions in the near term to attain a group goal. There is a tendency for more suggestions to be given by more people in computerized conferencing. This is part of the equalitarian tendency for more members to actively participate in the task behavior of a group in CC. In one of the problems, the difference was statistically significant at the .05 level; whereas in the other, it was sizable but did not reach statistical significance.

5. "Gives opinion, evaluation, analysis, expresses feeling, wish"

Includes all reasoning or expressions of evaluation or interpretation.

This is the most frequent type of communication for both problems and both modes. For the Bales problem, there was no difference in its prevalence associated with mode of communication. For the Arctic problem, however, there was a large and statistically significant difference, with more opinion giving in the CC condition.

6. "Gives Orientation, information, repeats, clarifies, confirms"

This includes statements that are meant to secure the attention of the other, (such as "There are two points I'd like to make..."), restating or reporting the essential content of what the group has read or said; non-inferential, descriptive generalizations or summaries of the situation facing the group. There are no clear differences here. Whereas there is a statistically significant difference in the direction of giving more orientation in CC for Forest Ranger, for the other problem, the difference is reversed.

7. "Asks for orientation, information, repetition and confirmation"

There is a significant tendency for this to occur more often in face to face discussions. This is probably because of the frequency with which a group member does not hear or understand the pronunciation of a sentence or partial utterance. In CC, people are usually more careful to state their thoughts clearly, and the recipient can read it several times rather than asking for repetition if it is not understood the first time or is later forgotten. We have noticed many CC participants going back and looking at comments a second or third time; in a face to face discussion, they would probably ask something like: "What was it you said before about x?".

8. "Asks for opinion, evaluation, analysis, expression of feeling"

This occurs more frequently in computerized conferencing. For one of the problems, the difference reached statistical significance, whereas it did not for the other. This tendency to more frequently and explicitly ask for the opinions of all the other group members, as well as to more spontaneously offer ones own opinions and analyses in CC, does seem to qualitatively be characteristic of the medium.

9. "Asks for suggestion, direction, possible ways of action"

This includes all overt, explicit requests, such as "What shall we do now?". It is not very prevalent in either medium, and there are no significant differences.

10. "Disagrees, shows passive rejection, formality, witholds resources"

This includes all the milder forms of uisagreement or refusal to comply or reciprocate. This is also an infrequent form of communication, but it occurs more in face to face discussions than in CC.

11. "Shows tension, asks for help, withdraws out of field"

Includes indications that the subject feels anxious or frustrated, with no particular other group member as the focus of these negative feelings. The results on this are rather puzzling. We end up with a statistically significant tendency for there to be more tensions when in CC for the Forest Ranger problem, but in FTF for the Arctic problem. Substantively, the proportion of these communications is very small in any case, and therefore, the small differences are not important.

12. "Shows antagonism, deflates other's status, defends or asserts self"

This includes autocratic attempts to control or direct others, rejection or refusal of a request, deriding or criticizing others.

This is infrequent in both media and there are no significant differences.

EFFECTS OF ORDER

For the most part, it did not matter whether the CC or the FtF discussion was held first. However, more suggestions were offered on the arctic problem if it was discussed in CC as the first problem, but more in FTF discussion if the FTF was preceeded by a CC condition. This is consistent with the tendency for CC to promote more giving of suggestions; apparently, the tendency carries over to a subsequent face to face conversation. This raises the interesting possibility that the group process and structure can be permanently changed by the experience of interacting through CC, a change that will carry over even to communications in other modes. Other pieces of evidence from other studies, including self reports of participants in long term field trials, indicate the same possibility.

CONCLUSION

Our investigation confirms the hypothesis that there are some significant differences in the group communication process between face to face and computer mediated discussions. Such differences seem to be associated with other characteristics of the medium, such as the greater tendency for minority opinions to be maintained, rather than a total group consensus emerging. In a fuller analysis (Hiltz, Johnson, Aronovitch and Turoff, 1980) we show that the observed differences in interaction profiles are highly correlated with the ability of a group to reach consensus and with the quality of group decision reached.

APPENDIX

Analyses of Variance
Bales Categories by Mode and Problem
2x2x4 Nested Factorial
Arctic
Individual % Data
Bales Category 1 - Shows Solidarity

Means
Mode of Communication

		FTF	CC	
Order of Problem	1st	1.6893	2.4348	2.0620
	2nd	1.6228	2.4437	2.0333
		1.6561	2.4392	

Nested Design

Source	SS	df	MS	F
A	12.2673	1	12.2673	3.9004
B	.0166	1	.0166	.0053
A x B	.0285	1	.0285	.0091
C/AB	37.7414	12	3.1451	1.3745
S/ABC	146.4430	64	2.2881	
Tot.	196.4967	79		

Table Values For F
1 and 12 df=4.75
12 and 64df=1.90

Pooled ANOVA

Source	SS	df	MS	F
A	12.2673	1	12.2673	5.0618*
B	.0166	1	.0166	.0068
A x B	.0285	1	.0285	.0117
WG	184.1844	76	2.4234	
Tot.	196.4967	79		

Table Value for F
1 and 76 df=3.97
*Significant

A = mode
B = order
C/AB = error term for AB, and A x B
S/ABC = error term for C/AB
WG = Pooled error term
The pooled design yields a significant difference between the FTF and CC conditions. The CC conditions show a greater percent of their comments in the category of shows solidarity.

2x2x4 Nested Factorial
Forest Ranger
Individual % Data
Bales Category 3 - Agrees

Means
Mode of Communication

		FTF	CC	
Order of Problem	1st	14.1900	5.4645	9.8273
	2nd	12.1921	4.1183	8.1552
		13.1910	4.7914	

Nested Design

Source	SS	df	MS	F
A	1411.0740	1	1411.0740	32.8693*
B	55.9134	1	55.9134	1.3024
A x B	2.1232	1	2.1232	.0495
C/ABC	515.1580	12	42.9298	.6774
S/ABC	4056.1449	64	63.3772	
Tot.	6040.4135	79		

Table Values for F
1 and 12 df=4.75
12 and 64 df=1.90
*Significant

Pooled ANOVA

The following pooled design is not really necessary since one finds the variables significant as above.

Source	SS	df	MS	F
A	1411.0740	1	1411.0740	23.4598*
B	55.9134	1	55.9134	.9296
A x B	2.1232	1	2.1232	.0353
WG	4571.3029	76	60.1487	
Tot.	6040.4135	79		

Table Value for F
1 and 76 df=3.97
*Significant

A=mode
B=order
C/AB=error term for A, B, A x B
S/ABC=error term for C/AB
WG=Pooled error term

The nested design yields a significant difference between the FTF and CC Conditions. The FTF conditions show a greater percent of their comments in category 3-Agrees.

REFERENCES

Bales, Robert
1950 Interaction Process Analysis; A Method for the Study of Small Groups. Reading, Mass; Addison Wesley.

Bales, Robert F. and Edgar F. Borgatta
1955 "Size of Group as a Factor in the Interaction Profile." In A.P. Hare, E. F. Borgatta and R. F. Bales, eds., Small Groups: Studies in Social Interaction, pp. 396-413. New York: Knopf.

Eady, Patrick M. and J. Clayton Lafferty
1975 "The Subarctic Survival Situation." Plymouth, Michigan: Experiential Learning Methods.

Hiltz, Starr Roxanne
1975 "Communications and Group Decision Making"; Experimental Evidence on the Potential Impact of Computer Conferencing. Newark, N.J., Computerized Conferencing and Communications Center, New Jersey Institute of Technology, Research Report No. 2.

Hiltz, Starr Roxanne, Kenneth Johnson, Charles Aronovitch and Murray Turoff
1980 Face to Face Vs. Computerized Conferences: A Controlled Experiment.

Hiltz, Starr Roxanne, Kenneth Johnson, and Ann Marie Rabke
1980 Communications Process and Outcome in Face to Face Vs. Computerized Conferences.

Hiltz, Starr Roxanne and Murray Turoff
1978 The Network Nation: Human Communication via Computer. Reading, Mass.: Addison Wesley Advanced Book Program.

ACKNOWLEDGEMENTS

The research reported here is supported by a grant from the Division of Mathematical and Computer Sciences (MCS 78-00519). The findings and opinions reported are solely those of the authors, and do not necessarily represent those of the National Science Foundation.

Murray Turoff and Charles Aronovitch played a large part in the design and analysis for this project. We are also grateful to Julian Scher and Peter and Trudy Johnson-Lenz for their contributions to the design of the experiments; to John Howell and James Whitescarver for their software design and programming support; and to our research assistants for their dedicated efforts in carrying out the experiments and coding questionnaires: Joanne Garofalo, Keith Anderson, Christine Naegle, Ned O'Donnell, Dorothy Preston, Stacy Simon and Karen Winters.

We would also like to thank Robert Bales and Experimental Learning Methods for their cooperation in providing documentation and permission to use adaptations of problem solving tasks which they originally developed.

WHAT TYPE OF INTERACTION IS IT TO BE

Emanuel A. Schegloff
Department of Sociology, U.C.L.A.

For one, like myself, who knows something about human interaction, but next to nothing about computers and human/machine interaction, the most useful role at a meeting such as this is to listen, to hear the troubles of those who work actively in the area, and to respond when some problem comes up for whose solution the practices of human interactants seems relevant. Here, therefore, I will merely mention some areas in which such exchanges may be useful.

There appear to be two sorts of status for machine/technology under consideration here. In one, the interactants themselves are humans, but the interaction between them is carried by some technology. We have had the telephone for about 100 years now, and letter writing much longer, so there is a history here; to it are to be added video technology, as in some of the work reported by John Carey, or computers, as in the "computer conferencing" work reported by Hiltz and her colleagues, among others. In the other sort of concern, one or more of the participants in an interaction is to be a computer. Here the issues seem to be: should this participant be designed to approximate a human interactant? What is required to do this? Is what is required possible?

1) If we take as a tentative starting point that person-person interaction should tell us what machine-person interaction should be like (as Jerry Hobbs suggests in a useful orienting set of questions he circulated to us), we still need to determine what type of person-person interaction we should consult. It is common to suppose that ordinary conversation is, or should be, the model. But that is but one of a number of "speech-exchange systems" persons use to organize interaction, or to be organized by in it."Meetings," "debates," "interviews," and "ceremonies" are vernacular names for other technically specifiable, speech-exchange systems orgainzing person-person interaction. Different types of turn-taking organization are involved in each, and differences in turn-taking organization can have extensive ramifications for the conduct of the interaction, and the sorts of capacities required of the interactants. In the design of computer interactants, and in the introduction of technological intermediaries in human-human interaction, the issue remains <u>which type</u> of person-person interaction is aimed for or achieved. For example, in the Pennsylvania video link-up of senior citizen homes, John Carey asks whether the results look more like conversation or like commercial television. But many of details he reports suggests that the form of technological intervention has made what resulted most like a "meeting" speech exchange system.

2) The term "interactive" in "interactive program" or in "person/machine interaction" seems to refer to no more than that provision is made for participation by more than one participant. "Interactive" in this sense is not necessarily "interactional," i.e., the determination of at least some aspects of each party's participation by collaboration of the parties. For the "talk" part of person-person interaction, a/the major vehicle for this "interactionality" is the sequential organization of the talk; that is, the construction of units of participation with specific respect to the details of what has preceded, and thereby the sequential position in which a current bit of talk is being done. Included among the relevant aspects of "what has preceded" and "current sequential position" is "temporality," or "real time," though not necessarily measured by conventional chronometry. What are, by commonsense standards, quite tiny bits of silence -- two tenths of a second, or less (what we call micro-pauses) -- can, and regularly do, have substantial sequential and interactional consequences. The character of the talk after them is regularly different, or is subject to different analysis, interpretation or inference.

Although the telephone deprives interactants of visual access to each other, it leaves this "real time" temporality largely unaffected, and with it the integrity of sequential organization. Nearly all the technological interventions I have heard about -- whether replacing an interactant, or inserted as [illegible] medium between interactants -- impacts on this aspect [illegible] e exchange of talk. It is one reason for wondering [illegible] r retention of ordinary conversation as the tar[illegible] this enterprise is appropriate. For some of the [illegible] mplated innovations, like computer conferencing, exchanges of letters may be a more appropriate past model to study, for there too more than one may "speak" at a time, long lapses may intervene between messages, sequential ordering may be puzzling (as in "Did the letters cross in the mail?") etc.

3) Sequential organization has a direct bearing on an issue which must be of continuing concern to workers in this area -- that of understanding and misunderstanding. It is the sequential (including temporal) organization of the talk which, in ordinary conversation, provides running evidence to participants that, and how, they have been understood. The devices by which troubles of understanding are addressed (what we call "repair," discussed for computers by Phil Hayes in a recent paper) -- requests for repetition or clarification and the like -- are only one part of the machinery which is at work. Regularly, in ordinary conversation, a speaker can detect from the produced-to-be-responsive next turn of another s/he has or has been, misunderstood, and can immediately intervene to set matters right. This is a major safeguard of "intersubjectivity," a retention of a sense that the "same thing" is being understood as what is being spoken of. The requirements on interactants to make this work are substantial, but in ordinary conversation, much of the work is carried as a by-product of ordinary sequential organization. The anecodotes I have heard about misunderstandings going undetected for long stretches when computers are the medium, and leading to, or past, the verge of nastiness, suggest that these are real problems to be faced.

4) In all the business of person-person interaction there operates what we call "recipient-design" -- the design of the participation by each party by reference to the features (personal and idiosyncratic, or categorial) of the recipient or co-participant. The formal machineries of turn-taking, sequential organization, repair, etc. are always conditioned in their realization on particular occasions and moments by this consideration. I don't know how this enters into plans for computerized interactants, and it remains to be seen how it will enter into the participation of humans dealing with computers. Persons make all sorts of allowances for children, non-native speakers, animals, the handicapped, etc. But there are other allowances they do not make, indeed that don't present themselves as allowances or allowables. What is involved here is a determination of where the robustness is and where the brittleness, in interacting with persons by computers, for in the areas of robustness it may be that many of the issues I've mentioned may be safely ignored; the people "will understand."

Throughout these notes, we are at a very general level of discourse. The real pay-offs, however, will come from discussing specifics. For that, interaction will be needed, rather than position papers.

THE COMPUTER AS AN ACTIVE COMMUNICATION MEDIUM

John C Thomas
IBM T J Watson Research Center
PO Box 218 Yorktown Heights, New York 10598

1. THE NATURE OF COMMUNICATION

Communication is often conceived of in basically the following terms. A person has some idea which he or she wants to communicate to a second person. The first person translates that idea into some symbol system which is transmitted through some medium to the receiver. The receiver receives the transmission and translates it into some internal idea. Communication, in this view, is considered good to the extent that there is an isomorphism between the idea in the head of the sender before sending the message and the idea in the receiver's head after recieving the message. A good medium of communication, in this view, is one that adds minimal noise to the signal. Messages are considered good partly to the extent that they are unabmiguous. This is, by and large, the view of many of the people concerned with computers and communication.

For a moment, consider a quite different view of communication. In this view, communication is basically a design-interpretation process. One person has goals that they believe can be aided by communicating. The person therefore designs a message which is intended to facillitate those goals. In most cases, the goal includes changing some cognitive structure in one or more other people's minds. Each receiver of a message however has his or her own goals in mind and a model of the world (including a model of the sender) and interprets the received message in light of that other world information and relative to the perceived goals of the sender. This view has been articulated further elsewhere [1].

This view originates primarily from putting the rules of language and the basic nature of human beings in perspective. The basic nature of human beings is that we are living organisms and our behavior is goals-directed. The rules of language are convenient but secondary. We can language rules for a purpose break

Communicating in different media produces different behaviors and reactions [2, 3]. The interesting first order finding however, is that people *can* communicate using practically any medium that lets any signal through if motivation is high enough. We can, under some circumstances, communicate with people who use different accents, grammars, or even languages. Yet, in other circumstances, people who are ostensibly friends working on a common goal and who have known each other for years end up shouting at each other: 'You're not listening to me. No, you don't understand!'

One fundamental aspect of human communication then is that it is terrifically adaptive, and robust,containing a number of sophisticated mechanisms such as explanations that simultaneously facillitate social and work goals [4], metacomments that direct the conversation [5], and rules for taking turns [6].

To the extent that these mechanisms can be embedded in a computer system that is to dialogue with humans, the dialogue will likely tend to be more successful. However, equally true of human communication is that it is sometimes quite ineffective. Let us examine where, why, and how the computer can help improve communication in those cases.

2. FUNDAMENTAL DIFFICULTIES IN COMMUNICATION

The view of communication as a design-interpretation process suggests that since messages are designed and interpretted to achieve goals, the perceived relationship between the goals of the communicators is likely to be a powerful determinant of what happens in communication. Common observation as well experimental results [1] are consistent with this notion. People often view themselves in situations of pure competition or pure cooperation. In fact, I suggest that either perception is due to a limited frame. Any two people who view themselves as involved in a zero-sum game are doing so because they have a limited frame of reference. In the widest possible frame of reference, there is at least one state probabilistically influenced by their acts (such as the total destruction of human life through nuclear weapons) that both would find undesirable. Therefore, when I am playing tennis, poker, or politics with someone and we say we are in pure competition, we are only doing so in a limited framework. In a wider framework, it is always in our mutual interest to cooperate under certain circumstances.

This does not mean, however, that people perceive this wider framework. Because of the limitations of human working memory, people often forget that there is a framework in which they can cooperate. Indeed, this describes one of the chief situations in which a so-called breakdown of communications occurs. If we are *truly* in a zero-sum game, communication is only useful to the extent that we mislead, threaten, etc.

Conversely, people are only in pure cooperation by limiting their framework. I suggest that it is highly likely, given any two individuals, that they would put a different preference ordering on the set of all possible states of the world which their actions could probabilistically affect. This gives rise to a second type of breakdown in communication. People appear to be desiring to cooperate but they are only cooperating with respect to some limited framework X. They are competing with respect to some larger framework X plus Y. The most common X plus Y is X, the framework of cooperation plus Y, a consideration of whose habits must change for mutually beneficial action in the framework X.

For instance, two tennis partners obviously both want to win the game. Yet one is used to playing with both partners attempting to take the net. The other is used to the 'one-up, one-back' strategy. They can get into a real argument. What they are competing about is basically who is going to change, whose opinion is wrong, and similar issues. This then, in a sense, is a second type of breakdown of communication.

A third case exists even within the framework of cooperation. This case of difficult communication exists when the presupposed conceptual frameworks of the communicators is vitally discrepant. A computer programmer really wants to help a business person automate his or her invoicing application and the business person really wants this to happen. However, each party erroneously presumes more shared knowledge and viewpoint than in fact exists.

A puzzle still remains however. If people have such sophisticated, graceful, robust communication mechanisms, why do they not quite readily and spontaneously overcome these communication blocks?

WIDESPREAD ANTI-PRODUCTIVE BELIEFS

The biggest stumbling blocks to effective communication are the individual communicator's beliefs. People *typically* hold beliefs which are not empirically based. To some extent, it is impossible not to. In order to simplify the world sufficiently to deal with it, we make generalizations. If it turns out on closer inspection that these genralizations are correct, we call it insight while if it turns out that they are incorrect, we call it overgeneralization.

There are, however, a number of specific non-empirically based beliefs that people are particularly likely to believe which are anti-productive to communication. Among these are the following: 1. I must be understood; 2. If the other person disagrees with me, they don't understand me; 3. My worth is equal to my performance; 4. Things should be easy; 5. The world must be fair; 6. If I have the feeling of knowing something is true, it must be true; 7. If the other person thinks my idea is wrong, the person thinks little of me; 8. If this person's idea is wrong, the person is worthless; 9. I don't need to change -- they do; 10. Since I already know I'm right, it is a waste of time to really try to see things from the other person's perspective. 11. If I comprehend something, in the sense that I can rephrase it in a syntactically different way, that means I have processed deeply enough what the other person is saying. 12. I must tell the truth at all times no matter what. 13. If they cannot put it in the form of an equation (or computer program, or complete sentences, or English), they don't really know what they are talking about and so it is not possibly in my interest to listen.

Each of the above statements, has a correlated, less rigid, less extreme statement that is empirically based. For instance, if we really thought 'When I am wrong, some people will temporarily value me less', that is a valid generalization. In contrast, the thought 'When I am wrong, people will value me less' is an overgeneralization.

Similarly, it is quite reasonable to believe that expressing something mathematically has advantages and that if it is not expressed mathematically it may be more difficult for me to use the ideas; it may even be so difficult that I choose not to bother. It is not empirically based to believe that it is never worth you while to attempt to understand things not expressed in equations.

Nearly everyone, even quite psychotic people hold rational as well as irrational beliefs. Very few people when asked whether they have to be perfect in everything will say yes. However, very many people reject so completely evidence that they may be fundamentally wrong, that they act as though they must be perfect. It is bitter irony that most people can think and feel much more clearly about the things that are less important to them such as a crossword puzzle than they can about things that are much more important such as their major decisions in work and love.

Now let us imagine someone who has done a certain office procedure a certain way for many years. Then someone begins to explain a new procedure that is claimed to work better. There are a number of wholly rational reasons why the experienced office worker can be skeptical. But it is probably quite worthwhile to at least attempt to really understand the other person's ideas before criticizing them. There are many non-empirically based beliefs that may interfer in the communication process. The experienced office worker may, for instance, notice the young age of the systems analyst and believe that no-one so young could really understand what is going on. They may believe that if there is a better way, they should have seen it themselves years ago and if they didn't they must be an idiot. Since they didn't see it and they can't be an idiot, there must not be a better way. They may just think to themselves it will be too hard to learn a new way. Very effective individual therapy [7] is based on trying to identify and change an individual's irrational beliefs. The focus of this paper however is on how a computer system could aid communication by overcoming or circumventing such irrational beliefs in those cases where communication appears to break down.

We know that people are capable of changing from a narrow competition framework to a wider cooperative framework in order to communicate. People can resolve differences about whose behavior needs to change. Normal communication has the mechanisms to do these things; when they fail to happen it is often because of irrational beliefs which prevent people from attempting to see things from the other person's perspective.

The tennis partner's disagreeing about what strategy to use will tend to resolve the disagreement without detriment to their mutual goal of winning the game, provided their thinking stays fairly close to the empirical level. If, however, one of the participants finds a

flaw in the other's thinking and then overgeneralizes and thinks 'What an idiot. That doesn't logically follow. How can anyone be so dumb.' But by the token 'dumb', the angry person probably means 'all-around bad.' Now this is an extrememly counter-productive overgeneralization which will tend to color the person's thinking on other issues of the game which are not even within the scope of the argument about what strategy to use. In extremely irrational but not so uncommon cases, the person may even express to the other person verbally or non-verbally that they have a generally low opinion of their partner. If either party becomes angry, they are also likely to mix up their messages about their own internal state with messages about the content of the game. Thus, 'I am angry.' gets mixed with 'A serve to that person's backhand will probably produce a weaker return.' The result may be a statement like 'Why can't you serve to his backhand for a change.' Such a statement is likely to increase the probability of serves to the forehand or double faults to the backhand.

Once each person becomes angry with the other, they are almost certainly overgeneralizing to the extent that they are believing that the only way to improve the situation is for the other person to change their behavior in some way 'He should apologize to me for being such an idiot.' No active problem solving behavior remains directed where it belongs: 'How can I improve the situation myself? How can I communicate better?' This is communication breakdown.

4. *THE POSSIBLE USES OF AN ACTIVE COMMUNICATION CHANNEL*

Now, let's just for the sake of arguement, *assume* or if you like *pretend* that what I have said so far is a useful perspective. What about the computer? In particular, what about using the power of the computer as a non-transparent ACTIVE medium of communication? The computer has been very successfully used as a way for people to communicate which allows speed/repetition and demands precision. Is there also a way for the computer to be used to enhance party-to-party communication in a way that helps defeat or get around the self-defeating beliefs that get in the way of effective communication in situations where participants have similar goals but are working in different frameworks? Can the computer aid in situations where participants have partially similar goals but are concentrating on the differences...or are unable to arrive at conclusions that are in both parties self-interest because of interferrence from a set of separate issues where they are in fundamental conflict?

An entire technology equal to the one that has addressed the speed/repetion precision issues could be built around this task. Clearly I cannot provide this technology myself in fifteen minutes or fifteen years. But let me provide one example of the *kind* of thing I mean. Suppose that one two people were disagreeing and communicating via Visual Display Terminals connected to a computer network. Let us suppose that the computer network imposed a formalism on the communication. Suppose, for example that strength and directionality of current emotional state were encoded on a spatially separate channel from content messages. Imagine that the designer of the message had to choose what emotion or emotions they felt and attempt to honestly quantify these. This information would be presented to the other person separately from the content statements. One unfortunate human weakness would be overcome; viz., the tendency to let the emotional statement -- 'I am angry' intrude into the content of what is said.

Now, suppose the computer network presented to the interpretter of this message a set of signals labelled as follows: 'The person sending this message to you is currently producing the following emotional states in themselves: Anger +7, Anxiety +4, Hurt +3, Depression +2, Gladness -6.' Note that the attribution has also been shifted squarely to where it belongs -- on the person with the emotional state.

Now suppose further that when a person stated their position, certain key words triggered a request by the system for restatement. For instance, suppose a person typed in 'You always get what you want.' The system may respond with: 'Regarding the word 'always', could you be more quantitative. First, in how many instances during the last two weeks would you estimate that there have been occassions when that person would like to have gotten something but could not get that thing?'

Unfortunately, asked just such a question, an angry person would probably become angrier and direct some anger toward the active channel itself. A marriage counselor is often caught in just this sort of bind, but can usually avoid escalating anger via empathy and other natural mechanisms. How a computerized system could avoid increasing anger remains a challenge.

Another possibility would be for the channel to enforce the protocol for conflict resolution suggested by Rappaport and others. For instance, before stating your position, you would have to restate your opponent's position to their satisfaction.

Needless to say, participants using such an active interface would be apprized of the fact and voluntarily choose to use such an interface for their anticipated mutual benefit in the same way that labor and management often agree to use a mediator or arbitrator to help them reach an equitable solution. Unfortunately, such a choice requires that both the people involved recognize that they are not perfect -- that their communication ability could use an active channel. This in itself presupposes some dismissal of the erroneous belief that their worth EQUALS their performance. Most people are capable of doing this before they become emotionally upset and hence might well agree ahead of time to using such a channel.

5. *SUMMARY*

In this paper, I reiterate the view that for many purposes, communication is best conceived of as a

design-interpretation process rather than a sender-receiver process. Fundamental difficulties in two-person communication occur in certain common situations. The incidence, exacerbation, and failure to solve such communication problems by the parties themselves can largely be traced to the high frequency of strongly held anti-empirical belief systems. Finally, it is suggested that the computer is a medium for humans to communicate with each other VIA. Viewed in this way, possibilities exist for the computer to become an *active* and *selective* rather than a *passive, transparent* medium. This could aid humans in overcoming or circumventing communication blocking irrational beliefs in order to facillitate cooperative problem solving.

6 REFERENCES

[1] Thomas, J. A Design-Interpretion Analysis of Natural English. *International Journal of Man Machine Studies*, 1978, *10*, 651-668.

[2] Carey, J. A Primer on Interactive Television, *Journal of the University Film Association*, 1978, *XXX* (2), 35-39.

[3] Chapanis, A. Interactive Human Communication: Some Lessons Learned from Laboratory Experiments. Paper presented at NATO Advanced Study Institute on 'Man-Computer Interaction', Mati, Greece, 1976.

[4] Wynn, E. Office Conversation as an Information Medium. (In preparation).

[5] Thomas, J. A Method for Studying Natural Language Dialogue, *IBM Research Report*, 1976, RC-5882.

[6] Sacks, H., Schlegloff, E., and Jefferson, G. A Simplest Systematics for the Organization of Turn-taking for Conversation, *Language*, 1974, *50* (4), 696-735.

[7] Ellis, A. *Reason and Emotion in Psychotherapy*. New York: Lyle Stuart, (1962).

WHAT DISCOURSE FEATURES AREN'T NEEDED IN ON-LINE DIALOGUE

Eleanor Wynn
Xerox Office Products Division
Palo Alto, California

It is very interesting as a social observer to track the development of computer scientists involved in AI and natural language-related research in theoretical issues of mutual concern to computer science and the social study of language use. The necessity of writing programs that demonstrate the validity or invalidity of conceptualizations and assumptions has caused computer scientists to cover a lot of theoretical ground in a very short time, or at least to arrive at a problem area, and to see the problem fairly clearly, that is very contemporary in social theory. There is in fact a discrepancy between the level of sophistication exhibited in locating the problem area (forced by the specific constraints of programming work) and in the theorizations concocted to solve the problem. Thus we find computer scientists and students of language use from several disciplines converging in their interest in the mechanics and metaphysics of social interaction and specifically its linguistic realization. Attempts to write natural language programs delivered the realization that even so basic a feature as nominal reference is no simple thing. In order to give an "understander" the wherewithal to answer simple questions about a text, one had to provide it with an organized world in which assumptions are inferred, in which exchanges are treated as part of a coherent and minimally redundant text, in which things allow for certain actions and relations and not others, and for which it is unclear how to store the information about the world in such a way that it is accessible for all its possible purposes and delivered up in an appropriate way. Some of these were providable and some weren't. Some AI workers have already moved into the phenomenological perspective, just from confronting these problems -- a long way to go from the assumptions of mathematics, science, and engineering that they originally brought to the task.

Others, in their attempts to deal with issues of representation and motivation in discourse, have started recreating segments of the history of social theory. This is the history and perspective that students of social interaction bring with them to the problem. They arrive at the problem area either through a theoretical evolutionary process in which they reject the previous stage of theory, and interaction is a good demonstration of the limitations of that theory, or because they are simply intrigued by observing the wealth of social action with which they can identify as members, that the study of naturally-occuring discourse provides.

In social theory, the ethnomethodological perspective arose as a response to the:

1) political implications
2) reifications
3) unexamined assumptions
4) narrow filter on observation

presented by structural-functionalist theory.

This theory:

1) limits and constructs observation fairly strictly
2) justifies the status quo (whatever exists serves a survival function)
3) posits a macro-organization (well-defined institutions and roles)
4) uses platonic idealizations of the social order
5) is normative
6) doesn't explain change very well

[illegible]ficulties in this theory were in part an artifact of a general positivist-scientistic orientation in which there was a motivation to treat the social world as a scientific object and hence to structure the descriptions of it in such a way as to make the social world amenable to prediction, testing and control. The ethnomethodological or phenomenological perspective does not give up the scientific pretension but it does drop the engineering [illegible]. A world whose modes of [illegible] and saying [illegible] or practices are constantly [illegible] created or [illegible] which, though following [illegible] recognizable tracks, is in a constant state of invention and confirmation, lends itself far less to prediction. In fact it is clearly unpredictable.

Language itself provides an analogy, though it is partly the character of language that allows for the constant state of invention in the social world. Language changes constantly by means of several mechanisms, among which are phonological drift, usage requirements, metaphorization, and social emulation based on values and fashions. For theoretical purposes, one of the most valuable findings in Labov's landmark quantitative studies of phonological variation, was that social values drive the distribution of optional variants from one speech occasion to another according to the perceived formality of the occasion. In this manner, values -- what individuals at different social levels consider to be prestigious articulations, drive phonological change in general. Linguistic fashions themselves also change in response to what is currently used, and change with or against the majority according to the kind of identification desired to be made. They cannot be predicted in advance as such changes in value are typically discovered not planned. Very often changes in language use are derivative, based on a secondary or marginal meaning or usage, or discovered analogy or metaphor of some existing locution. Thus a dynamic of social contrasts and identifications, as well as social mobility and aspirations thereto, as well as socially situated invention, are deeply connected to linguistic issues, including language change and the concept of distribution rules, in an empirically observable and countable way. These and other social dynamics operate no less for more complex discourse phenomena, and account for large portions of observed discourse strategies.

Generally, when a sociolinguist, sociologist, or anthropologist looks at language use, what they attend to are the disclosed social practices. Being aware of, and focussing on social context, with a history of social theory or an historically developed set of concepts for social action in mind, alerts one to many attributes of the occasion for interaction: the possible social identities and relationships of the participants, the perceived outcomes and the social significance of meanings generated in the course of the interaction, as well as to structural and habitual features that reflect social requirements (viz. the "recognition" requirement as a prerequisite to interaction's taking place at all or in the particular form, as discussed by Schegloff).

The fact that a background of shared knowledge about the world is assumed emerges from an examination of what is explicitly stated and from the observation that what is explicit is in some way "incomplete", partial, not a full itemization of what is communicated and understood. It is also the case that to spell out all the assumptions would be unbearably time-consuming, redundant to the purpose, boring, and possibly an infinite regress; and this practice would moreover fail to accomplish all those conversational purposes which require negotiation, building up to a point of mutual orientation and accord, or the "use" of one person by another for a real or imaginary gain. (cf Simmel)

The messiness, potential ambiguity, implicitness, etc. of natural conversation serve many of the purposes that actors have, including the one of intimacy and mutuality by less and less explicit surface discourse. Herein lies an important distinction, one that is not well perceived by workers in AI. Purposes can be, and typically are <u>discovered</u> in the course of interaction rather than planned. Purposes are thus emergent from interaction rather than apriori organizing principles of it.

Attempting to code, catalogue, regulate, formalize, make explicit in advance those purposes is reminiscent of structuralist, positivist social theory. To this extent, computer scientists are recreating social theory, starting from the point that is most amenable to their hopes and needs, and so far lacking the dialectic that contextualizes other developments in social theory. Ontogeny has not yet fully recapitulated phylogeny. Extending the plans, goals, frames notion into the wider social world (wider than a story understander), constitutes a platonic idealization and the ensuing problem of locating those idealizations somewhere, as if there were large programs running in our heads (some of which need debugging), or as if there were some accessible pool of norms from which we draw each time we act. It posits that we act out these idealizations in our everyday behavior, that our behavior constitutes realized instances of this structure. This conflicts with a "process" notion of interaction, which careful discourse analysis reveals, whereby participants are continually trying out and signalling their participation in a mutual world, presumably because this is not from one instance to the next pre-given. The great revelation of discourse analysis in general, if I may be so sweeping, is the ability to observe the process of social action, whereby the social world is essentially built up anew for the purpose at hand, and interactants can be seen sorting out the agreed-on premises from those that need to be established between them.

There are two kinds of concerns here that bear upon on-line dialogue research. One is the notion of person, social identity, etc. The other is the notion of interaction as a reality testing mechanism that grounds the individual in a chosen point of view from among the many interpretations available to him for any given "event". Both of these notions differentiate the computer from a person as an interactant. Sorting out dialogue issues that embody these notions, narrows down the field of concerns that are relevant for building "robust" on-line dialogue systems.

All social systems, including non-human ones, display social differentiation. This is a central notion that the AI path of evolution does not bring to the study of discourse. On the contrary, discourse problems are treated as if there were a universality among potential interactants. This fits very nicely with a platonic perspective. Kling and Scacchi have referred to this as the rationalist perspective, and they cite claims made for simulation and modelling as their illustration of how exponents of this perspective fail to make even gross social distinctions:

"Neglecting the obiter dicta claim that modelling and simulation are 'applicable to essentially all problem-solving and decision-making,' presumably including ethical decisions, one is left with an odd account of the problem of modelling. Models are 'far from ubiquitous' and 'the trouble is' they are difficult and costly to develop and use. But the appropriateness of modelling is not linked by (rational perspectivists) to any discernible social setting or the interests of its participants. (Their) claims are not aimed at policy-making in particular. They could include simulations

for engineering design as well as for projecting the costs of new urban development. However, their comments typify the rational perspective when it is applied to information systems in policy-making; the presumption is that differences in social settings make no difference."

Work in socio-linguistics, on the other hand, has focussed on how speech varies by situation, by relationship, by purpose and by many other constraints that depend upon both a typification of the other from a complex set of loose attributes and the discovery of his unique behavior in the situation. The notion of a linguistic "repertoire" expresses people's demonstrated ability and propensity to adjust their speech at almost every analytic level, down to the phonology, to their perception of the situation and the audience. There are variations in people's skill at this, but all do it. To the extent that they don't do it, they risk being inappropriate and not getting rewards from interaction. (see F. Erickson for a study of the outcomes of interactive strategies in ethnically mixed interactions.)

The structuralist perspective again may be an appealing way for computer scientists to approach the problem of differentiation of persons, as it posits an essentially limited set of "roles" of fairly fixed attributes, and posits as well an ordered hierarchical arrangement of those roles. With this framework in mind it is relatively easier to imagine a computer as a viable participant in a social interaction, as it should be possible to construct an identifiable role for it. With this rather flat view of human social perception it is also possible to imagine a person requiring of a computer that it behave appropriately in a conversation, without regard for the fact that a computer can only satisfy a very limited set of purposes for that person in interaction. In fact people know perfectly well many of the things computers can't do for them or to them, things which other people can do and hence which need to be taken into account in dealing with other people. And they are able to differentiate for the purpose of interaction among infinitely many people, and states of mind or situation those people can be in.

The other feature of interaction between people, reality-testing, is less well understood than differentiation, which is a veritable solid ground of social understanding. However, it can be seen in interactions, even very simple task-oriented ones such as I described in my thesis, that people are also always accessing each other for a view of the world, for agreement, disagreement, and a framework for interpreting. Diffuse explanation mechanisms(Wynn, 1979) also exhibit the tendency of speaker to nail down the audience's perception of himself to the framework of interpretation desired by him, as an implicit acknowledgement of possible variance. What is often uncertain in an actor's "model" or projection, or understanding of the other participants or observers, is their view of the actor himself. To this end, he fills in and guides the interpretation with additional context any time he perceives an occasion for misinterpretation, sometimes to the point of logical absurdity (but practical appropriateness if not necessity).

Since a computer is not an actor in the social world, its interpretations, both of oneself and of "events" -- perceived social phenomena-- don't really count. A computer can provide facts about the world within a well-understood framework, but it cannot provide the kind of context that comes from being a participant in social life, nor a validation of another's perception, except to the extent that matters of "fact" or true-false distinctions allow this. And in these cases, the person supplies this validation himself from the information. This may be a moot point, but I maintain that the search for agreement, confirmation, etc., and the related search for common ground or reality are basic motives for interaction, along with confirmations of membership and solidarity etc., as described in the work of Schegloff and of much earlier writers like Malinowski and Simmel.

Rather than working from careful and detailed observations of the real world, excepting such innovators as Grosz and Robinson, many computer scientists exhibit a tendency to develop their "models" of interaction by conceptualizing from the perspective of the machine and its capabilities or possible capabilities. Discourse features may be selected for attention and speculation because they offer either a machine analog or a machine contrast. Thus we people are attributed information structures, search procedures and other constructs which are handy metaphors from the realm of computerdom; and it would be especially handy if we were in fact constructed according to these clean notions, so that our thinking and behavior could be modelled. (In all fairness, I know computers have "guys" running around inside them, "going" places, "looking for" stuff, trying out things, getting excited or upset, going nuts, giving up, etc.)

Working from the machine perspective can lead to some gross observational oversights, and the authors of the oversight I've picked as an example will hopefully indulge me. The implicit confirmationhypothesis (Hayes and Reddy) could never have been hypothesized by anyone who studies language behavior from a social perspective, as one of the oldest conversational observations around is the explicit confirmation observation. The phatic communion notion is over 30 years old, and is perhaps the first attention given to those features of interaction whichwere initially considered to carry little or no observable propositional content or information. Included in these behaviorsare those discourse "fillers" that signal to the speaker he is being received with no problem, that the listener is still paying attention (even more basic than confirming), and that the listener is a participant in the rhythm of the interaction even though he is producing little speech at the moment. The "rights" and "hehhehheh's" of the current natural conversation transcription conventions are absolutely pervasive and omnipresent. Nods, "hm's", gaze, prompt questions, frowns, smiles, exclamations of wonder, are all explicit confirmation devices constantly used in conversation, and occur especiallywhennew propositions or details essential to building a story are presented. Speakers are also often tentative and reformulate at any evidence of withheld confirmation, like a "blank stare" or a frown from the audience.

Therefore it is by no means ungraceful to explicitly confirm, and on the other hand, it takes very little to do so. But the point is this: even if the implicit confirmation hypothesis were true (and I pick it because it is an available example and very easy to reject-- other notions would do as well but require a more detailed attack), it would be no reason to exclude this feature from a computer dialogue nor to suppose that it would pose people any difficulty in handling a dialogue with a machine. The discourse supporting activities of natural conversation always address practical concerns. If a new concern should arisebecause of newconstraints-- e.g. that the interactant is a machine--these will be incorporated in the ongoing details of communication. For instance, when it is obvious someone is having difficulty speakin- and understanding English, we unhesitatingly drop all ellipsis and give full articulation of every sound, even though this produces great redundancy in the message for purposes of communicating with another native speaker, and is moreover extremely unhabitual.

In fact, the social role of the computer is perhaps most like that of a foreigner. We assume a foreign individual whose English is poor to have an ability to communicate, perhaps a rudimentary grammar and vocabulary of our language, and a set of customs, some of which overlap with ours. But we can't take the specifics of any of these things for granted. There is very little in the way of a background of practices or assumptions to work with. But here the analogy ends.

Presumably, we won't be going to on-line dialogue programs to chit-chat. The purposes will be fairly well-defined and circumscribed. People will interact with a computer:

1) because there is no person available
2) because there is limited social confront in accessing expert information from a computer, so it is available in a metaphorical sense
3) because the computer has specialized abilities and resources not found in a single individual
4) because it coordinates non- local information and
5) is maximally up-to-date -- changes in status and the news of this are concurrently available and
6) the outcome of one's own interaction with the system may be animmediately registered action, like reserving a space and hence making one less space available to subsequent users
7) because actual searching (as opposed to the metaphoric kind attributed to our minds by cognitive scientists) of a large database may be required and the computer is much better and faster at this than we are.

In other words, our reasons, certainly our most solid and fulfillable reasons, for consulting computersand engaging in discourse with them will beto find out things relating to a framework we already have. The computer needs to know a few things about us and especially our language, and especially needs to know how to ask usto clarify what we said, even to present menus of intentions for us to choose from as a response to something unexecutable by it. But more than anything, it needs to be able to make its structure of information clear to us. In this sense it will satisfy certain "person" properties -- we have working notions of at least the parameters and starting points for negotiation with people. Whereas with computers we have at best an entry strategy for an unfamiliar system, but very little to go on in common knowledge for assessing its informedness or even consistency.

So on-line dialogue should not be like person-to-person dialogue in many respects. For instance, being overly explicit with a person is an indication of a judgment we have made about their competence. This judgment is quite likely to be offensive if it's wrong. (Schegloff) This is not likely to be a problem with a computer from an experiential social action point of view. Who cares if the computer cannot perceive that we are competent members of some social category defined bya more or less common body of knowledge: We will have no problem in telling it what level to address in dealing with us, if it has any such levels of explicitness, nor in gearing our own remarks to the appropriate level once we find out what it can digest. On-line dialogue systems therefore have an ongoing task of representing themselves, not the whole interactive world; and designers need not concern themselves so much with providing their systems with models of users, but rather providing users with clear models of the system they are interacting with. These are the major concerns, obviously.

I wish I could now deliver the part of the paper that would be of most interest: what a dialogue system should contain and how it can make available those contents in order to realize the purposes just stated. Instead I have addressed myself to what look like common fallacies that I see in attempting to incorporate natural language dialogue issues into computer dialogue issues without access to the social understandings embedded in social interaction research.

Interactive Discourse: Looking to the Future
Panel Chair's Introduction

Bonnie Lynn Webber
University of Pennsylvania

In any technological field, both short-term and long-term research can be aided by considering where that technology might be ten, twenty, fifty years down the pike. In the field of natural language interactive systems, a 21 year vision is particularly apt to consider, since it brings us to the year 2001. One well-known vision [1] of 2001 includes the famous computer named Hal - one offspring, so to speak, of the major theoretical and engineering breakthrough in computers that Clarke records as having occurred in the early 1980's. This computer Hal is able to understand and converse in perfect idiomatic English (written and spoken) with the crew of the spacecraft Discovery. And not just task-oriented dialogues, mind you!

Hal is a far cry from today's prototype natural language query systems, intelligent CAI-systems, diagnostic assistance systems, and Kurzweil machines. For one thing, Hal is not just responsive: he takes the initiative. His first documented utterance on board the spacecraft Discovery comes at a time when the crewmen Bowman and Poole are engrossed in a fading vision screen image of Poole's family on Earth, on the occasion of Poole's birthday.

> "Sorrv to interrupt the festivities," said Hal, "but we have a problem."

Not only can Hal converse in perfect idiomatic English, but he is a master of problem context (Panel 1) and social context (Panel 2) as well!

Now Hal is clearly where we currently are not at, and 2001 is clearly only one man's vision (albeit a very special man). Yet Clarke's depiction of Hal raises several issues, which along with other ones, provide a cue for the current panel discussion. The issues include:

1. Where is it that we want to have, must have, can expect to have, or conversely, should not have to have, Natural Language Interactive Systems?

2. Barring Clarke's reliance on the triumph of automatic neural network generation, what are the major hurdles that still need to be overcome before Natural Language Interactive Systems become practical?

3. What effects can we expect, deriving from the availability of, what to me seem, almost magical developments in hardware?

4. Are there practical (and acceptable) alternatives to interacting with machines in natural language in the various situations that provide a positive answer to question 1?

5. Should we be shooting for spoken Natural Language interactions - either input or output or both - or should we not, like Clarke, go the whole way and expect our machines to read lips as well.

REFERENCES

1. Clarke, Arthur C., 2001: A Space Odyssey, New American Library, 1968.

PROSPECTS FOR PRACTICAL NATURAL LANGUAGE SYSTEMS

Larry R. Harris
Artificial Intelligence Corporation
Newton Centre, Mass. 02159

As the author of a "practical" NL data base query system, one of the suggested topics for this panel is of particular interest to me. The issue of what hurdles remain before NL systems become practical strikes particulary close to home. As someone with a more pragmatic view of NL processing, my feeling is, not surprisingly, that we already have the capability to construct practical NL systems. Significant enhancement of existing man-machine communication is possible within the current NL technology if we set our sights appropriately and are willing to take the additional effort to craft systems actually worthy of being used. The missing link isn't a utopian parsing algorithm yet to be discovered. The hurdles to practical NL systems are of a much more conventional variety that require, as Edison said, more perspiration than inspiration.

It should be clear that none of my remarks conflict with the obvious fact that NL research has miles to go and that there are innumerable unresolved issues that will continue to require research beyond the foreseeable future. Our understanding of NL has merely scratched the surface, and it is fair to say that we don't even understand what all the problems are, muchless their solution. But by using the powerful techniques that have already resulted from NL research in extremely restricted micro-worlds it is possible to attain a high enough level of performance to be of practical value to a significant user community. It is these highly specialized systems that can be made practical using the existing technology.

I will not speculate on when a general NL capability will become practical, nor will I speculate on whether the creation of practical specialized systems will contribute to the creation of a more general capability. The fact that there is a clear need for improved man-machine communication and that current specialized systems can be built to meet that need, is reason enough to construct them.

The issue of whether practical specialized NL systems can now be built is, in my opinion, not a debatable issue. Those of us on this panel and other researchers in the field, simply don't have the right to determine whether a system is practical. Only the users of such a system can make that determination. Only a user can decide whether the NL capability constitutes sufficient added value to be deemed practical. Only a user can decide if the system's frequency of inappropriate response is sufficiently low to be deemed practical. Only a user can decide whether the overall NL interaction, taken in toto, offers enough benefits over alternative formal interactions to be deemed practical.

If we accept my point that practicality is in the eyes of the user, then we are led to the inescapable conclusion that practical NL systems can now be built, because several commercial users of such a system [Pruitt, O'Donnel] have gone on record stating that the NL capability within the confines of data base query is of significant practical value in their environment. These statements plus the fact that a substantial body of users employ NL data base query in daily productive use clearly meets the spirit of a "practical" NL system.

The main point of my remarks is not to debate the semantics of practicality, but to point out that whatever level of utility has been achieved, is due only in small part to the sophistication of the NL component. The utility comes primarily from a custom fitting of the NL component to the exact requirements of the domain; and from the painstaking crafting of the lexicon and grammar to achieve tha necessary density of linguistic coverage. In a sense, practicality is derived from a pragmatic approach that emphasizes proper performance on the vast bulk of rather uninteresting dialog, rather than focusing on the much smaller portion of intellectually challenging input. A NL system that is extrememly robust within well-defined limitations is far more practical than a system of greater sophistication that has large gaps in the coverage.

Attaining this required level of robustness and density of linguistic coverage is not necessarily as intellectually challenging as basic research, nor is it necessarily even worthy of publication. But let's not kid ourselves -- it is absolutely necessary to achieve a practical capability! It has never been clear to me that members of the ACL were interested in practical NL systems, nor is it clear that they should be. But I think that it is fair to say that there aren't many practical NL systems because there aren't very many people trying to build them! I would estimate, on the basis of my experience, that it takes an absolute minimum of 2 years, and probably more like 3 years, to bring a successful research prototype NL system to the level of practicality. This "development" process is well known in virtually all scientific and engineering disciplines. It is only our naivete of software engineering that causes us to underestimate the magnitude of this process. I'm afraid the prospects for practical NL systems look bleak as long as we have many NL researchers and few NL developers.

Pruitt, J., "A user's experience with ROBOT," Proceedings of the Fourth Annual ADABAS User's Meeting, April, 1977.

O'Donnell, J., "Experience with ROBOT at DuPont," Natural Computer Conference Panel, May, 1980.

FUTURE PROSPECTS FOR COMPUTATIONAL LINGUISTICS

Gary G. Hendrix
SRI International

Preparation of this paper was supported by the Defense Advance Research Projects Agency under contract N00039-79-C-0118 with the Naval Electronic Systems Command. The views expressed are those of the author.

A. Introduction

For over two decades, researchers in artificial intelligence and computational linguistics have sought to discover principles that would allow computer systems to process natural languages such as English. This work has been pursued both to further the scientific goals of providing a framework for a computational theory of natural-language communication and to further the engineering goals of creating computer-based systems that can communicate with their human users in human terms. Although the goal of fluent machine-based nautral-language understanding remains elusive, considerable progress has been made and future prospects appear bright both for the advancement of the science and for its application to the creation of practical systems.

In particular, after 20 years of nurture in the academic nest, natural-language processing is beginning to test its wings in the commercial world [8]. By the end of the decade, natural-language systems are likely to be in widespread use, bringing computer resources to large numbers of non-computer specialists and bringing new credibility (and hopefully new levels of funding) to the research community.

B. Basis for Optimism

My optimism is based on an extrapolation of three major trends currently affecting the field:

(1) The emergence of an engineering/applications discipline within the computational-linguistics community.

(2) The continuing rapid development of new computing hardware coupled with the beginning of a movement from time-sharing to personal computers.

(3) A shift from syntax and semantics as the principle objects of study to the development of theories that cast language use in terms of a broader theory of goal-motivated behavior and that seek primarily to explain how a speaker's cognitive state motivates him to engage in an act of communication, how a speaker devises utterances with which to perform the act, and how acts of communication affect the cognitive states of hearers.

C. The Impact of Engineering

The emergence of an engineering discipline may strike many researchers in the field as being largely detached from the mainstream of current work. But I believe that, for better or worse, this discipline will have a major and continuing influence on our research community. The public at large tends, often unfairly, to view a science through the products and concrete results it produces, rather than through the mysteries of nature it reveals. Thus, the chemist is seen as the person who produces fertilizer, food coloring and nylon stockings; the biologist finds cures for diseases; and the physicist produces moon rockets, semiconductors, and nuclear power plants. What has computational linguistics produced that has affected the lives of individuals outside the limits of its own close-knit community? As long as the answer remains "virtually nothing," our work will generally be viewed as an ivory tower enterprise. As soon as the answer becomes a set of useful computer systems, we will be viewed as the people who produce such systems and who aspire to produce better ones.

My point here is that the commercial marketplace will tend to judge both our science and our engineering in terms of our existing or potential engineering products. This is, of course, rather unfair to the science; but I believe that it bodes well for our future. After all, most of the current sponsors of research on computational linguistics understand the scientific nature of the enterprise and are likely to continue their support even in the face of minor successes on the engineering front. The impact of an engineering arm can only add to our field's basis of support by bringing in new suport from the commercial sector.

One note of caution is appropriate, however. There is a real possibility that as commercial enterprises enter the natural-language field, they will seek to build in-house groups by attracting researchers from universities and nonprofit institutions. Although this would result in the creation of more jobs for computational linguists, it would also result in proprietary barriers being established between research groups. The net effect in the short term might actually be to retard scientific progress.

D. The State of Applied Work

1. Accessing Databases

Currently, the most commercially viable task for natural-language processing is that of providing access to databases. This is because databases are among the few types of symbolic knowledge representations that are computationally efficient, are in widespread use, and have a semantics that is well understood.

In the last few years, several systems, including LADDER [9], PLANES [29], REL [26], and ROBOT [8], have achieved relatively high levels of proficiency in this area when applied to particular databases. ROBOT has been introduced as a commercial product that runs on large, mainframe computers. A pilot REL product is currently under development that will run on a relatively large personal machine, the HP 9845. This system, or something very much like it, seems likely to reach the marketplace within the next two or three years. Should ROBOT- and REL-like systems prove to be commercial successes, other systems with increasing levels of sophistication are sure to follow.

2. Immediate Problems

A major obstacle currently limiting the commercial viability of natural-language access to databases is the problem of telling systems about the vocabulary, concepts and linguistic constructions associated with new databases. The most proficient of the application systems have been hand-tailored with extensive knowledge for accessing just ONE database. Some systems (e.g., ROBOT and REL) have achieved a

degree of transportability by using the database itself as a source of knowledge for guiding linguistic processes. However, the knowledge available in the database is generally rather limited. High-performance systems need access to information about the larger enterprise that provides the context in which the database is to be used.

As pointed out by Tennant [27], users who are given natural-language access to a database expect not only to retrieve information directly stored there, but also to compute "reasonable" derivative information. For example, if a database has the location of two ships, users will expect the system to be able to provide the distance between them--an item of information not directly recorded in the database, but easily computed from the existing data. In general, any system thatis to be widely accepted by users must not only provide access to database information, but must also enhance that primary information by providing procedures that calculate secondary attributes from the data actually stored. Data enhancement procedures are currently provided by LADDER and a few other hand-built systems. But work is needed to devise means for allowing system users to specify their own database enhancement functions and to couple their functions with the natural-language component.

Efforts are now underway (e.g. [26] [13]) to simplify the task of acquiring and coding the knowledge needed to transport high-performance systems from one database to another. It appears likely that soon much of this task can be automated or performed by a database administrator, rather than by a computational linquist. When this is achieved, natural-language access to data is likely to move rapidly into widespread use.

E. New Hardware

VLSI (Very Large Scale Integration of computer circuits on single chips) is revolutionizing the computer industry. Within the last year, new personal computer systems have been announced that, at relatively low cost, will provide throughputs rivaling that of the Digital Equipment KA-10, the time-sharing research machine of choice as recently as seven years ago. Although specifications for the new machines differ, a typical configuration will support a very large (32 bit) virtual address space, which is important for knowledge-intensive natural-language processing, and will provide approximately 20 megabytes of local storage, enough for a reasonable-size database.

Such machines will provide a great deal of personal computing power at costs that are initially not much greater than those for a single user's access to a time-shared system, and that are likely to fall rapidly. Hardware costs reductions will be particularly significant for the many small research groups that do not have enough demand to justify the purchase of a large, time-shared machine.

The new generation of machines will have the virtual address space and the speed needed to overcome many of the technical bottlenecks that have hampered research in the past. For example, researchers may be able to spend less time worrying about how to optimize inner loops or how to split large programs into multiple forks. The effort saved can be devoted to the problems of language research itself.

The new machines will also make it economical to bring considerable computing to people in all sectors of the economy, including government, the military, small business, and to smaller units within large businesses. Detached from the computer wizards that staff the batch processing center or the time-shared facility, users of the new personal machines will need to be more self reliant. Yet, as the use of personal computers spread, these users are likely to be increasingly less sophisticated about computation. Thus, there will be an increasing demand to make personal computers easier to use. As the price of computation drops (and the price of human labor continues to soar), the use of sophisticated means for interacting intelligently with a broad class of computer users will become more and more attractive and demands for natural-language interfaces are likely to mushroom.

F. Future Directions for Basic Research

1. The Research Base

Work on computational linguistics appears to be focusing on a rather different set of issues than those that received attention a few years ago. In particular, mechanisms for dealing with syntax and the literal propositional content of sentences have become fairly well understood, so that now there is increasing interest in the study of language as a component in a broader system of goal-motivated behavior. Within this framework, dialogue participation is not studied as a detached linguistic phenomenon, but as an activity of the total intellect, requiring close coordination between language-specific and general cognitive processing.

Several characteristics of the communicative use of language pose significant problems. Utterances are typically spare, omitting information easily inferred by the hearer from shared knowledge about the domain of discourse. Speakers depend on their hearers to use such knowledge together with the context of the preceding discourse to make partially specified ideas precise. In addition, the literal content of an utterance must be interpreted within the context of the beliefs, goals, and plans of the dialogue participants, so that a hearer can move beyond literal content to the intentions that lie behind the utterance. Furthermore, it is not sufficient to consider an utterance as being addressed to a single purpose; typically it serves multiple purposes: it highlights certain objects and relationships, conveys an attitude toward them, and provides links to previous utterances in addition to communicating some propositional content.

An examination of the current state of the art in natural-language processing systems reveals several deficiencies in the combination and coordination of language-specific and general-purpose reasoning capabilities. Although there are some systems that coordinate different kinds of language-specific capabilities [3] [12] [20] [16] [30] [17], and some that reason about limited action scenarios [21] [15] [19] [25] to arrive at an interpretation of what has been said, and others that attempt to account for some of the ways in which context affects meaning [7] [10] [18] [14], one or more of the following crucial limitations is evident in every natural-language processing system constructed to date:

Interpretation is literal (only propositional content is determined).

The user's knowledge and beliefs are assumed to be identical with the system's.

The user's plans and goals (especially as distinct from those of the system) are ignored.

Initial progress has been made in overcoming some of these limitations. Wilensky [28] has investigated the use of goals and plans in a computer system that interprets stories (see also [22] [4]). Allen and Perrault [1] and Cohen [6] have examined the interaction between beliefs and plans in task-oriented dialogues and have implemented a system that uses

information about what its "hearer" knows in order to plan and to recognize a limited set of speech acts (Searle [23] [24]). These efforts have demonstrated the viability of incorporating planning capabilities in a natural-language processing system, but more robust reasoning and planning capabilities are needed to approach the smooth integration of language-specific and general reasoning capabilities required for fluent communication in natural language.

2. Some Predictions

Basic research provides a leading indicator with which to predict new directions in applied science and engineering; but I know of no leading indicator for basic research itself. About the best we can do is to consider the current state of the art, seek to identify central problems, and predict that those problems will be the ones receiving the most attention.

The view of language use as an activity of the total intellect makes it clear that advances in computational linguistics will be closely tied to advances in research on general-purpose common-sense reasoning. Hobbs [11], for example, has argued that 10 seemingly different and fundamental problems of computational linguistics may all be reduced to problems of common-sense deduction, and Cohen's work clearly ties language to planning.

The problems of planning and reasoning are, of course, central problems for the whole of AI. But computational linguistics brings to these problems its own special requirements, such as the need to consider the beliefs, goals, and possible actions of multiple agents, and the need to precipitate the achievement of multiple goals through the performance of actions with multiple-faceted primary effects. There are similar needs in other applications, but nowhere do they arise more naturally than in human language.

In addition to a growing emphasis on general-purpose reasoning capabilities, I believe that the next few years will see an increased interest in natural-language generation, language acquisition, information-science applications, multimedia communication, and speech.

Generation: In comparison with interpretation, generation has received relatively little attention as a subject of study. One explanation is that computer systems have more control over output than input, and therefore have been able to rely on canned phrases for output. Whatever the reason for past neglect, it is clear that generation deserves increased attention. As computer systems acquire more complex knowledge bases, they will require better means of communicating their knowledge. More importantly, for a system to carry on a reasonable dialogue with a user, it must not only interpret inputs but also respond appropriately in context, generating responses that are custom tailored to the (assumed) needs and mental state of the user.

Hopefully, much of the same research that is needed on planning and reasoning to move beyond literal content in interpretation will provide a basis for sophisticated generation.

Acquisition: Another generally neglected area, at least computationally, is that of language acquisition. Berwick [2] has made an interesting start in this area with his work on the acquisition of grammar rules. Equally important is work on acquisition of new vocabulary, either through reasoning by analogy [5] or simply by being told new words [13]. Because language acquisition (particularly vocabulary acquisition) is essential for moving natural-language systems to new domains, I believe considerable resources are likely to be devoted to this problem and that therefore rapid progress will ensue.

Information Science: One of the greatest resources of our society is the wealth of knowledge recorded in natural-language texts; but there are major obstacles to placing relevant texts in the hands of those who need them. Even when texts are made available in machine-readable form, documents relevant to the solution of particular problems are notoriously difficult to locate. Although computational linguistics has no ready solution to the problems of information science, I believe that it is the only real source of hope, and that the future is likely to bring increased cooperation between workers in the two fields.

Multimedia Communication: The use of natural language is, of course, only one of several means of communication available to humans. In viewing language use from a broader framework of goal-directed activity, the use of other media and their possible interactions with language, with one another, and with general-purpose problem-solving facilities becomes increasingly important as a subject of study.

Many of the most central problems of computational linguistics come up in the use of any medium of communication. For example, one can easily imagine something like speech acts being performed through the use of pictures and gestures rather than through utterances in language. In fact, these types of communicative acts are what people use to communicate when they share no verbal language in common.

As computer systems with high-quality graphics displays, voice synthesizers, and other types of output devices come into widespread use, an interesting practical problem will be that of deciding what medium or mixture of media is most appropriate for presenting information to users under a given set of circumstances. I believe we can look forward to rapid progress on the use of multimedia communication, especially in mixtures of text and graphics (e.g., as in the use of a natural-language text to help explain a graphics display).

Spoken Input: In the long term, the greatest promise for a broad range of practical applications lies in accessing computers through (continuous) spoken language, rather than through typed input. Given its tremendous economic importance, I believe a major new attack on this problem is likely to be mounted before the end of the decade, but I would be uncomfortable predicting its outcome.

Although continuous speech input may be some years away, excellent possibilities currently exist for the creation of systems that combine discrete word recognition with practical natural-language processing. Such systems are well worth pursuing as an important interim step toward providing machines with fully natural communications abilities.

G. Problems of Technology Transfer

The expected progress in basic research over the next few years will, of course, eventually have considerable impact on the development of practical systems. Even in the near term, basic research is certain to produce many spinoffs that, in simplified form, will provide practical benefits for applied systems. But the problems of transferring scientific progress from the laboratory to the marketplace must not be underestimated. In particular, techniques that work well on carefully selected laboratory problems are often difficult to use on a large-scale basis. (Perhaps this is because of the standard scientific practice of selecting as a subject for experimentation the simplest problem exhibiting the phenomena of interest.)

As an example of this difficulty, consider knowledge representation. Currently, conventional database management systems (DBMSs) are the only systems in widespread use for storing symbolic information. The AI community, of course, has a number of methods for maintaining more sophisticated knowledge bases of, say, formulas in first-order logic. But their complexity and requirements for great amounts of computer resources (both memory and time) have prevented any such systems from becoming a commercially viable alternative to standard DBMSs.

I believe that systems that maintain models of the ongoing dialogue and the changing physical context (as in, for example, Grosz [7] and Robinson [19]) or that reason about the mental states of users will eventually become important in practical applications. But the computational requirements for such systems are so much greater than those of current applied systems that they will have little commercial viability for some time.

Fortunately, the linguistic coverage of several current systems appears to be adequate for many practical purposes, so commercialization need not wait for more advanced techniques to be transferred. On the other hand, applied systems currently are only barely up to their tasks, and therefore there is a need for an ongoing examination of basic research results to find ways of repackaging advanced techniques in cost-effective forms.

In general, the basic science and the application of computational linguistics should be pursued in parallel, with each aiding the other. Engineering can aid the science by anchoring it to actual needs and by pointing out new problems. Basic science can provide engineering with techniques that provide new opportunities for practical application.

REFERENCES

1. Allen, J. & C. Perrault. 1978. Participating in Dialogues: Understanding via plan deduction. Proceedings, Second National Conference, Canadian Society for Computational Studies of Intelligence, Toronto, Canada.

2. Berwick, R. C., 1980. Computational Analogues of Constraints on Grammars: A Model of Syntactic Acquisition. The 18th Annual Meeting of the Association for Computational Linguistics, Philadelphia, Pennsylvania, June 1980.

3. Bobrow, D. G., et al. 1977. GUS, A Frame Driven Dialog System. Artificial Intelligence, 8, 155-173.

4. Carbonell, J. G. 1978. Computer Models of Social and Political Reasoning. Ph.D. Thesis, Yale University, New Haven, Connecticut.

5. Carbonell, J. G. 1980. Metaphor--A Key to Extensible Semantic Analysis. The 18th Annual Meeting of the Association for Computational Linguistics, Philadelphia, Pennsylvania, June 1980.

6. Cohen, P. 1978. On knowing what to say: planning speech acts. Technical Report No. 118, Department of Computer Science, University of Toronto. January 1978.

7. Gross, B. J., 1978. Focusing in Dialog. Proceedings of TINLAP-2, Urbana, Illinois, 24-26 July, 1978.

8. L. R. Harris, 1977. User Oriented Data Base Query with the ROBOT Natural Language Query System. Proc. Third International Conference on Very Large Data Bases, Tokyo (October 1977).

9. G. G. Hendrix, E. D. Sacerdoti, D. Sagalowicz, and J. Slocum, 1978. Developing a Natural Language Interface to Complex Data. ACM Transactions on Database Systems, Vol. 3, No. 2 (June 1978).

10. Hobbs, J. 1979. Coherence and coreference. Cognitive Science. Vol. 3, No. 1, 67-90.

11. Hobbs, J. 1980. Selective inferencing. Third National Conference of Canadian Society for Computational Studies of Intelligence. Victoria, British Columbia. May 1980.

12. Landsbergen, S. P. J., 1976. Syntax and Formal Semantics of English in PHLIQA1. In Coling 76, Preprints of the 6th International Conference on Computational Linguistics, Ottawa, Ontario, Canada, 28 June - 2 July 1976. No. 21.

13. Lewis, w. H., and Hendrix, G. G., 1979. Machine Intelligence: Research and Applications -- First Semiannual Report. SRI International, Menlo Park, California, October 8, 1979.

14. Mann, W., J. Moore, & J. Levin 1977. A comprehension model for human dialogue. Proceedings, International Joint Conference on Artificial Intelligence, 77-87, Cambridge, Mass. August 1977.

15. Novak, G. 1977. Representations of knowledge in a program for solving physics problems. Proceedings, International Joint Conference on Artificial Intelligence, 286-291, Cambridge, Mass. August 1977.

16. Petrick, S. R. 1978. Automatic Syntactic and Semantic Analysis. In Proceedings of the Interdsciplainary Conference on Automated Text Processing (Bielefeld, German Federal Republic, 8-12 November 1976). Edited by J. Petofi and S. Allen. Reidel, Dordrecht, Holland.

17. Reddy, D. R., et al. 1977. Speech Understanding Systems: A Summary of Results of the Five-Year Research Effort. Department of Computer Science. Carnegie-Mellon University, Pittsburgh, Pennsylvania, August, 1977.

18. Rieger, C. 1975. Conceptual Overlays: A Mechanism for the Interpretation of Sentence Meaning in Context. Technical Report TR-354. Computer Science Department, University of Maryland, College Park, Maryland. February 1975.

19. Robinson, Ann E. The Interpretation of Verb Phrases in Dialogues. Technical Note 206, Artificial Intelligence Center, SRI International, Menlo Park, Ca., January 1980.

20. Sager, N. and R. Grishman. 1975. The Restriction Language for Computer Grammars. Communications of the ACM, 1975, 18, 390-400.

21. Schank, R. C., and Yale A.I. 1975. SAM--A Story Understander. Yale University, Department of Computer Science Research Report.

22. Schank, R. and R. Abelson. 1977. Scripts, plans, goals, and understanding. Hillsdale N.J.: Laurence Erlbaum Associates.

23. Searle, J. 1969. Speech acts: An essay in the philosophy of language. Cambridge, England: Cambridge University Press.

24. Searle, J 1975. Indirect speech acts. In P. Cole and J. Morgan (Eds.), Syntax and semantics, Vol. 3, 59-82. New York: Academic Press.

25. Sidner, C. L. 1979. A Computational Model of Co-Reference Comprehension in English. Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, Massachusetts.

26. F. B. Thompson and B. H. Thompson, 1975. Practical Natural Language Processing: The REL System as Prototype. In M. Rubinoff and M. C. Yovits, eds., Advances in Computers 13 (Academic Press, New York, 1975).

27. H. Tennant, "Experience with the Evaluation of Natural Language Question Answerers," &Proc. Sixth International Joint Conference on Artificial Intelligene&, Tokyo, Japan (August 1979).

28. Wilensky, R. 1978. "Understanding Goal-Based Stories." Yale University, New Haven, Connecticut. Ph.D. Thesis.

29. D. Waltz, "Natural Language Access to a Large Data Base: an Engineering Approach," Proc. 4th Internatioal Joint Conference on Artificial Intelligence, Tbilisi, USSR, pp. 868-872 (September 1975).

30. Woods, W. A., et al. 1976. Speech Understanding Systems: Final Report. BBN Report No. 3438, Bolt Beranek and Newman, Cambridge, Massachusetts.

NATURAL LANGUAGE INTERACTION WITH MACHINES: A PASSING FAD? OR THE WAY OF THE FUTURE?

A. Michael Noll
American Telephone and Telegraph Company
Basking Ridge, New Jersey 07920

People communicate primarily by two modes: acoustic -- the spoken word; and visual -- the written word. It is therefore natural that people would expect their communications with machines to likewise use these two modes.

To a considerable extent, speech is probably the most natural of the natural-language modes. Hence, a fascination exists with machines that respond to spoken commands with synthetic speech responses to create a natural-language interactive discourse. However, although vast amounts of research and development effort have been expended in the search for systems that understand human speech and respond with synthetic speech, the goal of the perfect system remains as elusive as ever. Systems for producing natural-sounding speech for large vocabularies with unrestricted grammatical structures and for recognizing spoken speech for large vocabularies with unlimited grammatical structures and any number of talkers are still beyond the state of linguistics and computer science and technology.

Given the problems in the speech domain, it is not surprising that most interactions between people and machines are in the visual mode frequently using alphanumeric keyboards as input and textual display as output. Such visual terminals are already in fairly widespread use in industry and are used for a variety of applications including computer programming, text editing, and data-base access.

The telephone allows <u>speech</u> telecommunications over distance between people. Future visual terminals for the home and businesses will allow <u>textual</u> telecommunications between people. These visual terminals could also be used to telecommunicate with machines in a way that is presently difficult using the telephone and speech.

Viewdata, or videotex, systems are promised soon for the home and will allow data-base access and transactions with machines and textual messages between people. Some viewdata systems use elaborate tree searches to reach the desired frame of information. Some people believe that tree searches will be "unnatural" for many users and some other more-natural language will be needed to search and access these data-base systems.

One conclusion is that the future will see more choices in mode for telecommunications between people and with machines. The choice of which alternate mode will probably be dependent upon the specific application. For example, textual messages might be both easier to enter by keyboard and to read on a CRT screen than speaking to a recording machine and listening to a recorded message. However, social chatting might be best over the telephone. However, arranging a date with a stranger might be less revealing if done in the textual mode. Considerable opportunities exist for basic research to explore the suitability of these alternate modes for different communications applications.

The fascination of technologists with speech-synthesis chips is about to result in a variety of stand-alone appliances that speak. Ovens that state when the roast is done, washing machines that call for the addition of fabric softeners, automobiles that inform the driver that the door is open, and many other applications will soon abound in the marketplace. In most of these applications, synthetic speech will substitute for a lamp or other form of visual display. The environment will be polluted with the noise of buzzy synthetic speech. Many of these applications will undoubtedly be little more than passing fads.

But in some circumstances synthetic speech will become the way of the future. One example would be synthetic-speech announcements of floors in an elevator thereby eliminating crooked necks!

Most of the preceding examples are very restricted in terms of the language used for the interaction with machines. The problem with unrestricted natural language for communication with machines is that no automatic way has yet been discovered to extract meaning in either the speech or textual mode. The textual mode does eliminate the need for acoustic analysis and hence has been more extensively used in most systems for restricted, specialized applications. However, even if either mode were equally near perfect, questions would still arise about user preference for one mode over the other.

Thus, in the end the future will be decided by the votes of consumers in the marketplace as they choose from the many options presented by technology. The shrewd enterpreneur will use consumer preference and needs to help illuminate in advance the desires and needs of the marketplace. Basic research in linguistics, human behaviour, natural language, and other ancillary fields will have an important role in developing solutions and in understanding people's needs and behaviour.

NATURAL VS. PRECISE CONCISE LANGUAGES FOR HUMAN OPERATION OF COMPUTERS: RESEARCH ISSUES AND EXPERIMENTAL APPROACHES

Ben Shneiderman, Department of Computer Science
University of Maryland, College Park, MD.

This paper raises concerns that natural language front ends for computer systems can limit a researcher's scope of thinking, yield inappropriately complex systems, and exaggerate public fear of computers. Alternative modes of computer use are suggested and the role of psychologically oriented controlled experimentation is emphasized. Research methods and recent experimental results are briefly reviewed.

1. INTRODUCTION

The capacity of sophisticated modern computers to manipulate and display symbols offers remarkable opportunities for natural language communication among people. Text editing systems are used to generate business or personal letters, scientific research papers, newspaper articles, or other textual data. Newer word processing, electronic mail, and computer teleconferencing systems are used to format, distribute, and share textual data. Traditional record keeping systems for payroll, credit verification, inventory, medical services, insurance, or student grades contain natural language/textual data. In these cases the computer is used as a communication medium between humans, which may involve intermediate stages where the computer is used as a tool for data manipulation. Humans enter the data in natural language form or with codes which represent pieces of text (part number instead of a description, course number instead of a title, etc.). The computer is used to store the data in an internal form incomprehensible to most humans, to make updates or transformations, and to output it in a form which humans can read easily. These systems should act in a comprehensible "tool-like" manner in which system responses satisfy user expectations.

Several researchers have commented on the importance of letting the user be in control [1], avoiding acausality [2], promoting the personal worth of the individual [3], and providing predictable behavior [4]. Practitioners have understood this principle as well: Jerome Ginsburg of the Equitable Life Assurance Society prepared an in-house set of guidelines which contained this powerful claim:

"Nothing can contribute more to satisfactory system performance than the conviction on the part of the terminal operators that they are in control of the system and not the system in control of them. Equally, nothing can be more damaging to satisfactory system operation, regardless of how well all other aspects of the implementation have been handled, than the operator's conviction that the terminal and thus the system are in control, have 'a mind of their own,' or are tugging against rather than observing the operator's wishes."

I believe that control over system function and predictable behavior promote the personal worth of the user, provide satisfaction, encourage competence, and stimulate confidence. Many successful systems adhere to these principles and offer terminal operators a useful tool or an effective communication media.

An idea which has attracted researchers is to have the computer take coded information (medical lab test values or check marks on medical history forms) and generate a natural language report which is easy to read, and which contains interpretations or suggestions for treatment. When the report is merely a simple textual replacement of the coded data, the system may be accepted by users, although the compact form of the coded data may still be preferable for frequent users. When the suggestions for treatment replace a human decision, the hazy boundary between computer as tool and computer as physician is crossed.

Other researchers are more direct in their attempt to create systems which simulate human behavior. These researchers may construct natural language front ends to their systems allowing terminal operators to use their own language for operating the computer. These researchers argue that most terminal operators prefer natural language because they are already familiar with it, and that it gives the terminal operator the greatest power and flexibility. After all, they argue, computers should be easy to use with no learning and computers should be designed to participate in dialogs using natural language. These sophisticated systems may use the natural language front ends for question-answering from databases, medical diagnosis, computer-assisted instruction, psychotherapy, complex decision making, or automatic programming.

2. DANGERS OF NATURAL LANGUAGE SYSTEMS

When computer systems leave users with the impression that the computer is thinking, making a decision, representing knowledge, maintaining beliefs, or understanding information I begin to worry about the future of computer science. I believe that it is counterproductive to work on systems which present the illusion that they are reproducing human capacities. Such an approach can limit the researcher's scope of thinking, may yield an inappropriately complex system, and potentially exaggerates the already present fear of computers in the general population.

2.1 NATURAL LANGUAGE LIMITS THE RESEARCHER'S SCOPE

In constructing computer systems which mimic rather than serve people, the developer may miss opportunities for applying the unique and powerful features of a computer: extreme speed, capacity to repeat tedious operations accurately, virtually unlimited storage for data, and distinctive input/output devices. Although the slow rate of human speech makes menu selection impractical, high speed computer displays make menu selection an appealing alternative. Joysticks, lightpens or the "mouse" are extremely rapid and accurate ways of selecting and moving graphic symbols or text on a display screen. Taking advantage of these and other computer-specific techniques will enable designers to create powerful tools without natural language commands. Building computer systems which behave like people do, is like building a plane to fly by flapping its wings. Once we get past the primitive imitation stage and understand the scientific basis of this new technology (more on how to do this later), the human imitation strategies will be merely museum pieces for the 21st century, joining the clockwork human imitations of the 18th century. Sooner or later we will have to accept the idea that computers are merely tools with no more intelligence than a wooden pencil. If researchers can free themselves of the human imitation game and begin to think about using computers for problem solving in novel ways, I believe that there will be an outpouring of dramatic innovation.

2.2 NATURAL LANGUAGE YIELDS INAPPROPRIATELY COMPLEX SYSTEMS

Constructing computer systems which present the illusion of human capacities may yield inappropriately complex systems. Natural language interaction with the tedious clarification dialog seems archaic and ponderous when compared with rapid, concise, and precise database manipulation facilities such as Query-by-example or commercial word processing systems. It's hard to understand why natural language systems seem appealing when contrasted with modern interactive mechanisms like high speed menu selection, light pen movement of icons, or special purpose interfaces which allow the user to directly manipulate their reality. Natural language systems must be complex enough to cope with user actions stemming from a poor definition of system capabilities.

Some users may have unrealistic expectations of what the computers can or should do. Rather than asking precise questions from a database system, a user may be tempted to ask how to improve profits, whether a defendant is guilty, or whether a military action should be taken. These questions involve complex ideas, value judgments, and human responsibility for which computers cannot and should not be relied upon in decision making.

Secondly, users may waste time and effort in querying the database about data which is not contained in the system. Codd [5] experienced this problem in his RENDEZVOUS system and labeled it "semantic overshoot." In command systems the user may spend excessive time in trying to determine if the system supports the operations they have in mind.

Thirdly, the ambiguity of natural language does not facilitate the formation of questions or commands. A precise and concise notation may actually help the user in thinking of relevant questions or effective commands. A small number of well defined operators may be more useful than ill-formed natural language statements, especially to novices. The ambiguity of natural language may also interfere with careful thinking about the data stored in the machine. An understanding of onto/into mappings, one-to-one/one-to-many/many-to-many relationships, set theory, boolean algebra, or predicate calculus and the proper notation may be of great assistance in formulating queries. Mathematicians (and musicians, chemists, knitters, etc.) have long relied on precise concise notations because they help in problem solving and human-to-human communication. Indeed, the syntax of precise concise query or command language may provide the cues for the semantics of intended operations. This dependence on syntax is strongest for naive users who can anchor novel semantic concepts to the syntax presented.

2.3 NATURAL LANGUAGE GENERATES MISTRUST, ANGER, FEAR AND ANXIETY

Using computer systems which attempt to behave like humans may be cute the first time they are tried, but the smile is short-lived. The friendly greeting at the start of some computer-assisted instruction systems, computer games, or automated bank tellers, quickly becomes an annoyance and, I believe, eventually leads to mistrust and anger. The user of an automated bank teller machine which starts with "Hello, how can I help you?" recognizes the deception and soon begins to wonder how else the bank is trying to deceive them. Customers want simple tools whose range of functions they understand. A more serious problem arises with systems which carry on a complete dialog in natural language and generate the image of a robot. Movie and television versions of such computers produce anxiety, alienation, and fear of computers taking over.

In the long run the public attitude towards computers will govern the future of acceptable research, development, and applications. Destruction of computer systems in the United States during the turbulent 1960's, and in France just recently (Newsweek April 28, 1980 — An underground group, the Committee for the Liquidation or Deterrence of Computers claimed responsibility for bombing Transportation Ministry computers and declared: "We are computer workers and therefore well placed to know the present and future dangers of computer systems. They are used to classify, control and to repress.") reveal the anger and fear that many people associate with computers. The movie producers take their ideas from research projects and the public reacts to common experiences with computers. Distortions or exaggerations may be made, but there is a legitimate basis to the public's anxiety.

One more note of concern before making some positive and constructive suggestions. It has often disturbed me that researchers in natural language usually build systems for someone else to use. If the idea is so good, why don't researchers build natural language systems for their own use. Why not entrust their taxes, home management, calendar/schedule, medical care, etc. to an expert system? Why not encode their knowledge about their own disipline in a knowledge representation language? If such systems are truly effective then the developers should be rushing to apply them to their own needs and further their professional career, financial status, or personal needs.

3. HUMAN FACTORS EXPERIMENTATION FOR DEVELOPING INTERACTIVE SYSTEMS

My work with psychologically oriented experiments over the past seven years has made a strong believer in the utility of empirical testing [6]. I believe that we can get past the my-language-is-better-than-your-language or my-system-is-more-natural-and-easier-to-use stage of computer science to a more rigorous and disciplined approach. Subjective, introspective judgments based on experience will always be necessary sources for new ideas, but controlled experiments can be extremely valuable in demonstrating the effectiveness of novel interactive mechanisms, programming language control structures, or new text editing features. Experimental testing requires careful statement of a hypothesis, choice of independent and dependent variables, selection and assignment of subjects, administration to minimize bias, statistical analysis, and assessment of the results. This approach can reveal mistaken assumptions, demonstrate generality, show the relative strength of effects, and provide evidence for a theory of human behavior which may suggest new research.

A natural strategy for evaluating the effectiveness of natural language facilities would be to define a task, such as retrieval of ship convoy information or solution of a computational problem, then provide subjects with either a natural language facility or an alternative mode such as a query language, simple programming language, set of commands, menu selection, etc. Training provided with the natural language system or the alternative would be a critical issue, itself the subject of study. Subjects would perform the task and be evaluated on the basis of accuracy or speed. In my own experience, I prefer to provide a fixed time interval and measure performance. Since inter-subject variability in task performance tends to be very large, within subjects (also called repeated measures) designs are effective. Subjects perform the task with each mode and the statistical tests compare scores in one mode against the other. To account for learning effects, the expectation that the second time the task is performed the subject does better, half the subjects begin with natural language, while half the subjects begin

with the alternative mode. This experimental design strategy is known as counterbalanced orderings.

If working systems are available, then an on-line experiment provides the most realistic environment, but problems with operating systems, text editors, sign-on procedures, system crashes, and other failures can bias the results. Experimenters may also be concerned about the slowness of some natural language systems on currently available computers as a biasing factor in such experiments. An alternative would be on-line experiments where a human plays the role of a natural language system. This appears to be viable alternative [7] if proper precautions are taken. Paper and pencil studies are a suprisingly useful approach and are valuable since administration is easy. Much can be learned about human thought processes and problem solving methods by contrasting natural language and proposed alternatives in paper and pensil studies. Subjects may be asked to write queries to a database of present a sequence of commands using natural language or some alternative mode [9].

There is a growing body of experiments that is helping to clarify issues and reveal problems about human performance with natural language usage on computers. Codd [5] and Woods [8] describe informal studies in user performance with their natural language systems. Small and Weldon [7] conducted the first rigorous comparison of natural language with a database query language. Twenty subjects worked with a subset of SEQUEL and an on-line simulated natural language system to composed queries. Shneiderman [9] describes a similar paper and pencil experiment comparing performance with natural language and a subset of SEQUEL. The results of both of these experiments suggest that precise concise database query language do aid the user in rapid formulation of more effective queries.

Damerau [10] reports on a field study in which a functioning natural language system, TQA, was installed in a city planning office. His system succeeded on 513 out of 788 queries during a one year period. Hershman, Kelly and Miller [11] describe a carefully controlled experiment in which ten naval officers used the LADDER natural language system after a ninety minute training period. In a simulated rescue attempt the system properly responded to 258 out of 336 queries.

Critics and supporters of natural language usage can all find heartening and disheartening evidence from these experimental reports. The contribution of these studies is in clarification of the research issues, development of the experimental methodology, and production of guidelines for developers of interactive systems. I believe that developers of natural language systems should avoid over-emphasizing their tool and more carefully analyze the problem to be solved as well as human capacities. If the goal is to provide an appealing interface for airline reservations, bank transactions, database retrieval, or mathematical problem solving, then the first step should be a detailed review of the possible data structures, control structures, problem decompositions, cognitive models that the user might apply, representation strategies, and importance of background knowledge. At the same time there should be a careful analysis of how the computer system can provide assistance by representing and displaying data in a useful format, providing guidance in choosing alternative strategies, offering effective messages at each stage (feedback on failures and successes), recording the history and current status of the problem solving process, and giving the user comprehensible and powerful commands.

Experimental research will be helpful in guiding developers of interactive systems and in evaluating the importance of the user's familiarity with:

1) the problem domain
2) the data in the computer
3) the available commands
4) typing skills
5) use of tools such as text editors
6) terminal hardware such as light pens, special purpose keyboards or unusual display mechanisms
7) background knowledge such as boolean algebra, predicate calculus, set theory, etc.
8) the specific system - what kind of experience effect or learning curve is there

Experiments are useful because of their precision, narrow focus, and replicability. Each experiment may be a minor contribution, but, with all its weaknesses, it is more reliable than the anecdotal reports from biased sources. Each experimental result, like a small tile in a mosaic which has a clear shape and color, adds to our image of human performance in the use of computer systems.

4. REFERENCES

1) Cheriton, D.R., Man-Machine interface design for time-sharing systems, Proceedings of the ACM National Conference, (1976), 362-380.

2) Gaines, Brian R. and Peter V. Facey, Some experience in interactive system development and application, Proceedings of the IEEE, 63, 6, (June 1975), 894-911.

3) Pew, R.W. and A.M. Rollins, Dialog Specification Procedure, Bolt Beranek and Newman, Report No. 3129, Revised Edition, Cambridge, Massachusetts, 02138, (1975).

4) Hansen, W.J., User engineering principles for interactive systems, Proceedings of the Fall Joint Computer Conference, 39, AFIPS Press, Montvale, New Jersey, (1971), 523-532.

5) Codd, E.F., HOW ABOUT RECENTLY? (English dialogue with relational databases using RENDEZVOUS Version 1), In B. Shneiderman (Ed.), Databases: Improving Usability and Responsiveness, Academic Press, New York, (1978), 3-28.

6) Shneiderman, B., Software Psychology: Human Factors in Computer and Information Systems, Winthrop Publishers, Cambridge, MA (1980).

7) Small, D.W. and L.J. Weldon, The efficiency of retrieving information from computers using natural and structured query languages, Science Applications Incorporated. Report SAI-78-655-WA, Arlington,Va., (Sept. 1977).

8) Woods, W.A., Progress in natural language understanding - an application to lunar geology, Proceedings of the National Computer Conference, 42, AFIPS Press, Montvale, New Jersey, (1973), 441-450.

9) Shneiderman, B., Improving the human factors aspect of database interactions, ACM Transactions on Database Systems, 3, 4, (December 1978a), 417-439.

10) Damerau, Fred J., The Transformational Query Answering System (TQA) operational statistics - 1978, IBM T.J. Watson Research Center RC 7739, Yorktown Heights, N.Y. (June 1979).

11) Hershman, R.L., R.T. Kelly and H.G. Miller, User performance with a natural language query system for command control, Navy Personnel Research and Development Center Technical Report 79-7, San Diego,CA, (1979).

NATURAL LANGUAGE AND COMPUTER INTERFACE DESIGN

MURRAY TUROFF
DEPARTMENT OF COMPUTER AND INFORMATION SCIENCE
NEW JERSEY INSTITUTE OF TECHNOLOGY

SOME ICONOCLASTIC ASSERTIONS

Considering the problems we have in communicating with other humans using natural language, it is not clear that we want to recreate these problems in dealing with the computer. While there is some evidence that natural language is useful in communications among humans, there is also considerable evidence that it is neither perfect nor ideal. Natural language is wordy (redundant) and imprecise. Most human groups who have a need to communicate quickly and accurately tend to develop a rather well specified subset of natural language that is highly coded and precise in nature. Pilots and police are good examples of this. Even working groups within a field or discipline tend over time to develop a jargon that minimizes the effort of communication and clarifies shared precise meanings.

It is not clear that there is any group of humans or applications for computers that would be better served in the long run by natural language interfaces. One could provide such an interface for the purpose of acclimating a group or individual to a computer or information system environment but over the long run it would be highly inefficient for a human to continue to use such an interface and would in a real sense be a disservice to the user. Those retrieval systems that allow natural language like queries tend to also allow the user to discover with practice the embedded interface that allows very terse and concise requests to be made of the system. Take the general example of COBOL, which was designed as a language to input business oriented programs into a computer that could be understood by non-computer types. We find that if we don't demand that programmers follow certain standards to make this possible, they will make their programs cryptic to the point where it is not understandable to anyone but other programmers.

It is interesting to observe that successful interfaces between persons and machines tend to be based upon one or the other of the two extreme choices one can make in designing a language. One is small, well defined vocabularies from which one can build rather long and complex expressions and the other is large vocabularies with short expressions. In some sense, "natural language" is the result of a compromise between these two opposing extremes. If we had some better understanding of the cognitive dynamics that shape and evolve natural language, perhaps the one useful natural language interface that might be developed would allow individuals and groups to shape their own personalized interface to a computer or information system. I am quite sure that given such a powerful capability, what a group of users would end up with would be very far from a natural language.

The argument is sometimes made that a natural language interface might be useful for those who are linguistically disadvantaged. It might allow very young children or deaf persons to better utilize the computer. I see it as immoral to provide a natural language introduction to computers to people who might mistakenly come to think of a computer as they would another human being. I would much prefer such individuals to be introduced to the computer with an interface that will give them some appreciation for the nature of the machine. For example, a very simple CAI language called PILOT has been used to teach grammar school children how to write simple lessons for their classmates. The ability of the young children to write simple question answer sequences and then see them executed as if the computer was able to use natural language is, I believe, far more beneficial to the child than giving him canned lessons as his or her first impression of what a computer is like.

COMPUTERIZED CONFERENCING

Since 1973 at the New Jersey Institute of Technology, we have been developing and evaluating the use of a computer as a direct aid to facilitating human communication. The basic idea is to use the processing and logical capabilities of the computer to aid in the communication and exchange of written text (Hiltz & Turoff, 1978). As part of this program we have been operating the Electronic Information Exchange System (EIES) as a source of field trial data and as a laboratory for controlled experimentation. Currently, EIES has approximately 600 active users internationally. Our current rate of operation is about 5,000 user hours a month; 8,000 messages, conference comments and notebook pages written a month and about 35,000 delivered each month. The average message is about 10 lines of text and the average comment or page is about 20 lines of text.

EIES offers the user a complete set of differing interfaces including menus, commands, self-defined commands and self programming of interfaces for individuals and groups. In addition to the standard message, conference and notebook features, EIES has been designed with the incorporation of a computer language called "INTERACT" that allows special communication structures and data structures to be integrated into the application of any specific group. Much of this capability has evolved since 1976 through a numerous set of alternative feedback and evaluation mechanisms. Our users include scientists, engineers, managers, secretaries, teenagers, students, Cerebral Palsy children and 80 year old senior citizens. In all this experience we have yet to hear a direct request or even implicit desire for any sort of natural language like interface.

To the contrary, we have indirect empirical data that supports the premise that a natural language like interface would be a disadvantage. For the most part, the behavior of users on EIES is very sensitive to the degree of experience they have had with the system. However, there is one key parameter which is insensitive to the degree of experience or the rate of use of the system. This is the number of items a user receives when he or she sits down at the terminal to use the system. This number stays at around 7 plus or minus 2. This is obviously a prescriptive effect the system has on the user as they get into the habit of signing on often enough so that they will not have more than around 7 new text items waiting for them. Users who have been cut off for a long period by a broken terminal or a vacation that denies them access usually give out textual screams of "information overload" when they find tons of text items waiting for them. In a real sense, it is natural language that is generating this information overload for the user. Another pertinent observation is that each user has three unique identifiers; a full name, a short nickname, and a three digit number. Some users always use nicknames and some always use numbers to address their messages but I have yet to encounter anyone who uses full names on a regular basis.

AUTOMATED ABSTRACTING

Our observations do point to one application where the ability to process natural language would be a significant augmentation of the users of computerized conferencing systems. We have a large number of conferences that have been going on for over a year and which contain thousands of comments. While a person entering such an on-going discussion can, in principle, go back and read the entire transcript or do selective retrieval on subtopics, it would be far preferable to be able to generate automatic summaries of such large text files. Even for regular use, the ability to get automated summaries would significantly raise the threshold of information overload and allow users to increase their level of communication activity and the amount of information with which they can deal meaningfully.

The goal of being able to process natural language has always been a bit of a siren's call and has a certain note of purity about it. Those striving for it sometimes lose sight of the fact that an imperfect system may still be quite useful when the perfect system may be unobtainable for some time. One of the important problems well recognized in the computer field is teaching computers how to "forget" or eliminate garbage. A less well recognized problem is the one of teaching a computer how to "give up" gracefully and go to a human to get help. In other words, the natural language systems that may have significant payoff in the next decade are those that blend the best talents of man and machine into one working unit.

In the computerized conferencing environment, this means that a person requesting a summary of a long conference probably knows enough about the substance to guide the computer in the process and to tailor the summary to particular needs and interests. In computerized conferencing, the ultimate goal is "collective intelligence" and one hopes that the appropriate design of a communication structure will allow a group of humans to pool their intelligence into something greater than any of its parts. If there is an automated or artificial intelligence system, then providing that system as a tool to a group of humans as an integral part of their group communication structure, the resulting intelligence of the group should be greater than the automated system alone. I believe a similar observation holds for the processing of natural language. Too often those working in natural language seem to feel that integrating humans into the analysis process would be an impurity or contaminant. In fact, it may be the higher goal than mere automation.

WRITING STYLE

A related area with respect to computerized conferencing is the observation that the style of writing in this medium of communication differs from other uses of the written or spoken version of natural language. First of all, there is a strong tendency to be concise and to outline complex discussions. We can observe this directly in the field trials and also observe that users bring group pressure upon those who start to write verbose items or items off the subject of interest to the group. The mechanism most commonly employed is the anonymous message. Also, in our controlled experiments on human problem solving (Hiltz, et al, 1980) we have found that there is no difference in the quality of a solution reached in a face-to-face environment or in a computerized conferencing environment. However, we do observe that the computerized conferencing groups use approximately 60% fewer words to do just as good a job as the face-to-face groups. Using Bales Interaction Process Analyses (content analyses), we have also confirmed significant differences in the content of the communications.

New users go through a learning period in which it may take 10 to 20 hours to feel comfortable in writing in conferences. We feel this is due to the subconscious recognition that people write differently in this medium than in letters, memos or other forms of the written language. The majority of what a new user writes (95%) will be messages the first five hours of usage and it takes about 100 hours until 25% of their writings are in conferences. Also, it is about 100 hours before they feel comfortable in writing larger text items in notebooks. One other aspect in the style change is the incorporation of many non-verbal ques into written form (HA! HA!, for example). One cannot see the nod of the head or hear a gentle laugh. Another aspect of natural language processing that can aid users in this form of communications is help in overcoming learning curves of this sort by being able to process the text of a group and provide a comparative analysis to new members of a group so they can more quickly learn the style of the group and feel comfortable in communicating with the group. One can carry this farther and ask for abilities to deal in certain levels of emotion such as: I would like to make my statement sound more assertive.

CONCLUSION

I do believe that this form of human communication will become as widespread and as significant as the phone has been to our society. The future application of natural language processing really lies in this area; however, it is not in the interface to the computer that this future rests but rather on the ability of this field to provide humans direct aids in processing the text found in their communications. Perhaps the real subject to address is not the one with which this panel was titled but the problems of person-machine interface to natural language processing systems. Or, better yet, person-machine integration within natural language processing. The computer processing of natural language needs to become the tool of the writer, editor, translator and reader. It also has to aid us in improving our ability to communicate. Most organizations are run on communications and the lore that is contained in those communications. With the increasing use of computers as communication devices, the qualitative information upon which we depend becomes as available for processing as the quantitative has been.

Reference:

THE NETWORK NATION: Human Communication Via Computer, Starr Roxanne Hiltz and Murray Turoff, Addison-Wesley Advanced Book Program, 1978.

FACE TO FACE VS. COMPUTERIZED CONFERENCING: A controlled Experiment, Hiltz, Johnson, Aronovitch and Turoff, Report of the Computerized Conferencing and Communications Center, NJIT, January 1980.

DAT

FILM

4